The Importance of Social Relationships for Mental and Physical Health.

Social-Epidemiological Studies using Life Course and Machine Learning Approaches.

Jeppesen, Linda Ejlskov

DOI (link to publication from Publisher):
10.5278/vbn.phd.med.00112

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

THE IMPORTANCE OF SOCIAL RELATIONSHIPS FOR MENTAL AND PHYSICAL HEALTH

SOCIAL-EPIDEMIOLOGICAL STUDIES
USING LIFE COURSE AND MACHINE LEARNING APPROACHES

BY
LINDA EJLSKOV JEPPESEN

DISSERTATION SUBMITTED 2018

AALBORG UNIVERSITY
DENMARK

The Importance of Social Relationships for Mental and Physical Health.

Social-Epidemiological Studies using Life Course and Machine Learning Approaches

Ph.D. Dissertation

By Linda Ejlskov Jeppesen
Aalborg University

AALBORG UNIVERSITY
DENMARK

Dissertation was submitted June, 2018

Dissertation submitted:	June, 2018
PhD supervisor:	Associate Prof. Dr. Henrik Bøggild Aalborg University
Assistant PhD supervisors:	Prof. Dr. Christian Torp-Pedersen Aalborg University Ass. Prof. Claus Dalsgaard Hansen Aalborg University
PhD committee:	Professor, Dr.med. Mogens Lytken Larsen Aalborg University Head of Department, Professor Richard Peter Ulm University Associate Professor, PhD Ulla Toft Region H - The Capital Region of Denmark
PhD Series:	Faculty of Medicine, Aalborg University
Department:	Department of Health Science and Technology

ISSN (online): 2246-1302
ISBN (online): 978-87-7210-219-1

Published by:
Aalborg University Press
Langagervej 2
DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk

Printed in Denmark by Rosendahls, 2018

Acknowledgements

I wrote this dissertation during my studies at the Public Health and Epidemiology group, Dept. of Health Science and Technology, Aalborg University and during my research stay at the Medical Research Council Unit for Lifelong Health and Ageing, University College London, United Kingdom. I would like to thank these two institutions for giving me the opportunity to work on my projects in their research environments. I am tremendously grateful to Professor Christian Torp-Pedersen for funding my PhD study. Through him, I was given the opportunity to better myself within the research of social determinants of health and quantitative methodological tools. For this opportunity, I cannot thank him enough and am deeply indebted. I am also grateful for the funding from Aalborg University, Knud Højgaards Fond, Augustinus Fonden and Oticon Fonden. This funding provided me with the financial help I needed to have a five-month research stay abroad.

As I quote from John Donne at the beginning of the first chapter: no man is an island, entire of itself. So it is for health, wellbeing and - as I have learned in the last few years - for research. Thus, there are several people I would like to thank for their support, help, advice and contributions to my work. All errors that remain are, of course, my own.

Most importantly, I would like to thank my three supervisors Henrik Bøggild, Christian Torp-Pedersen and Claus D. Hansen. You have all three introduced me to aspects of research and uniquely contributed with your expertise in important ways. To all of you, thank you for your friendly, calm and supportive manner. Claus, thank you for encouraging and enabling me to pursue a research project on social relationships and health in collaboration with Christian and yourself during my graduate in Sociology. Without this opportunity, I would not -

with all probability - be here today. Your expertise within social theory, your insights to and experience with social epidemiological register-based research have been invaluable in my learning and my dissertation. Christian, as I have stated before, I cannot thank you enough for providing me with the opportunity to do a PhD. You have provided invaluable advice on pursuing a career within academia and have supported me without fault in my first steps towards this goal. Your and Henrik's knowledge of and guidance on the use of epidemiological methods, data management and health outcomes have greatly enhanced my learning outcome and the quality of the scientific studies. Henrik, besides the aspects already mentioned, I am very thankful for your support and guidance throughout these last few years. Your encouragement, advice and your ability to steer me back on track when I have run off in a new scientific direction have been invaluable. Besides my supervisors, I would like to thank all my co-authors for their advice and contributions. Thanks to my colleagues at the Public Health and Epidemiology group and the Unit of Epidemiology and Biostatistics for an interdisciplinary and inspiring atmosphere. A special thanks to my fellow PhD students; Steen, Mikkel, Line, Rikke, Torill and Christina.

I deeply appreciate my stay and all the people I met while being at the unit for Lifelong Health and Ageing, University College London. A stay that would not have been possible without Diana Kuh, who gave me the opportunity to visit her and her research unit. A special thanks to Mai Stafford who acted as my local supervisor and who took me under my wing. I could not have wished for a better stay, and the collaboration with Diana and Mai have tremendously heightened my skills as a researcher. I would also like the thank the entire unit for making my stay as pleasant as it were and especially the fellow PhDs and postdocs who made my visit as much a joyous as a learning experience.

To all my friends and family - thank you for your endless support and your understanding in the final phases of the PhD. And last but not least, I am grateful for the enormous support and encouragement from my partner Jesper. Unwaveringly, he has put up with my travels abroad and at home, my many late nights at the office and has helped me through frustration and by reviewing most of my work. Lastly, Jesper's passion and dedication to his research has been and continues to be a source of inspiration for me.

Linda Ejlskov, Aalborg University, June 2018

Outline

The next few pages present first the English and then the Danish summary of this dissertation.

In **Chapter 1**, I introduce the social epidemiological domain to which this dissertation belongs. I present the overall purpose and specific aims of this dissertation at the end of chapter 1. In **Chapter 2**, I describe the materials and methods used to fulfil the aims of the dissertation as well as a description of the data sources used. **Chapter 3** presents a brief description the main results and some further analyses I have conducted to investigate the aims described at the end of chapter 1. In **Chapter 4**, I discuss the methods, the validity of the studies and the potential sources of bias which are important in relation to interpreting the results. Chapter **5** discusses the results in relations to the aims. The conclusion of this dissertation in relation to the aims are presented in **Chapter 6**. **Chapter 7** brings the results into perspective and suggests topics for future research.

The scientific papers (Study I-IV) are presented in full in **appendix A, B, C and D**. Located directly after each paper are the supplementary materials to each study (**appendix ??, ??, ?? and ??**). These contain sensitivity analyses, further descriptions of covariates and operationalisations as well as the validity of the short UCLA-scale used in Study III and IV.

This dissertation is based on the following four papers:

Study I: Ejlskov, L., Mortensen, R. N., Overgaard, C., Christensen, L. R., Vardinghus-Nielsen, H., Kræmer, S. R., Wissenberg, M., Torp-Pedersen, C. & Hansen, C. D. (2014). Individual social capital and survival: a population study with 5-year follow-up. Published in **BMC public health**, 14(1), 1025.

Study II: Ejlskov, L., Bøggild, H., Hansen, C.D., Wulff, J., Hansen, SM., Lange, T., Gerds, T. & Torp-Pedersen, C. The effect of early-life and adult socioeconomic position on development of lifestyle related diseases. Under review at **European Journal of Public Health.**

Study III: Ejlskov, L., Wulff, J, Bøggild, H., Kuh, D & Stafford, M. (2017). Assessing the relative importance of correlates of loneliness in later life. Gaining insight using recursive partitioning. Published in **Aging and Mental Health**, 1-8.

Study IV: Ejlskov, L., Bøggild, H., Kuh, D. & Stafford, M. Social relationship adversities throughout the life course and risk of loneliness in later life. Revise and Resubmit in **Ageing and Society.**

English summary

Social connections are the cornerstone of a human being's life and health. From the moment we are made the seemingly individual choices and social circumstances of our parents as well as the social encounters throughout life in part shape who we become. In addition, these social factors are also key for the propensity to live a good, long and healthy life. This PhD-dissertation consists of four quantitative social-epidemiological studies. These utilize different conceptual approaches to measure qualitative and quantitative aspects of social relationships and their associations with several adverse physical and mental health outcomes. An added important contribution of this dissertation is the utilization and demonstration of statistical methods and conceptual approaches from the field of machine learning and life course epidemiology.

The dissertation is based on three data sources: 1) A representative study sample of the inhabitants of the North Region Denmark in 2007 merged with data from three different national registers (The Central Population Register, The Danish National Patient Register and The Danish Income Statistics Register) with 14 years follow-up on all-cause mortality, 2) a Danish national cohort consisting of all Danish Citizens born in 1961-1971 with follow-up between 14-24 years on cardiovascular disease, diabetes and pulmonary obstructive lung disease and 3) The National Survey of Health and Development, a representative British study sample born in 1946 that is the longest running birth cohort study in the world with loneliness at age 68 as the outcome.

Several main findings are worth highlighting. Overall, the findings in this dissertation suggest that the degree to which current quantity and quality of social relationships protect against ill health are dependent upon both life course experiences and gender. First, evidence differs for whether gender differences occur depending on the aspect of social relationships measured and the outcome investigated. For most aspects

of social relationships and most outcomes, the studies performed in this dissertation suggest limited or no evidence that the association between social relationship measures and health are buffered by gender. Thus, the results suggest that gender may be a buffering factor for some health outcomes but not others and for some aspects of social relationships and not others. Conceptualising current social relationships as a resource for the individual to improve and maintain good health, individual social capital as the only social relationship conceptualisation shows evidence of differential associations with all-cause mortality for men compared to women. The findings suggest that higher levels of trust and more frequent social contact are associated with a lower risk of all-cause mortality for women. In contrast, more frequent social contact is associated with a higher risk of all-cause mortality for men.

Second, extending the scope to a life course perspective - social relationship adversities experienced throughout the life course are associated with loneliness among older adults. The results suggest that childhood social relationship adversities may be a sensitive period and that the detrimental effect of social relationship adversities accumulates over the life course. A key finding in Study IV is that the extent to which lack of social contact exacerbates loneliness depends on social experiences in earlier life stages. Further, that a high current quality of relationships may mitigate the negative effect that earlier social relationship adversities exert on loneliness.

Third, turning to a marker of the social environment in early-life, the socioeconomic position of one's parents during childhood or early-life is associated with both cardiovascular disease, diabetes and pulmonary obstructive lung disease in adulthood. The results suggest that those having grown up in a lower socioeconomic position have a higher risk of developing these three diseases regardless of birth year and ethnicity. The associations are strongest for pulmonary obstructive lung disease and with similar effect sizes for both men and women. The effect seems to dilute over time.

Fourth, the associations between early-life socioeconomic position and the risk of developing these lifestyle-related diseases were all partly mediated by socioeconomic position in adulthood. The largest mediating effect was seen for pulmonary obstructive lung disease with approximately 67% of the association being mediated by adult socioeconomic position for both men and women and the lowest was observed for cardiovascular disease with approximately 50% being mediated by adult socioeconomic position.

Fifth, demonstrating how techniques from machine learning can be used in the health sciences, another key finding was that among 42 correlates of loneliness among older adults, both the quality and the quantity of social relationships are among the most important predictors of loneliness - a key mental and social well-being health outcome.

One cannot use one conceptualisation of social relationships or be looking at one time point to give an adequate representation of the importance of social relationships for health. Seemingly different aspects of social relationships tap into the same underlying features of social relationships. All aspects work in tandem to give an overview of the importance of social relationships for health. In this dissertation, I also demonstrated how data analytical advances in other fields might further social epidemiology by providing a means to assess the relative importance of many identified predictors, which may help develop the design and targeting of health interventions.

Dansk resumé

Sociale relationer er hjørnestenen i et menneskes liv og sundhed. Fra det øjeblik, vi bliver skabt, er vores - på overfladen - individuelle valg, vores forældres sociale forhold og sociale oplevelser gennem livet med til at forme, hvem vi bliver. Derudover er disse sociale faktorer med til at forme sandsynligheden for at leve et godt, langt og sundt liv. Denne ph.d.-afhandling består af fire kvantitative social-epidemiologiske studier. Studierne anvender forskellige tilgange til at måle kvantitative og kvalitative aspekter af sociale relationer og disses sammenhænge med forskellige fysiske og psykiske sundhedsudfald. Et vigtigt bidrag i denne afhandling er brugen og demonstrationen af statistiske metoder og konceptuelle tilgange fra maskinlæring og livsforløbsepidemiologi.

Denne afhandling baserer sig på tre datakilder: 1) En repræsentativ stikprøve fra borgere i Region Nordjylland koblet med data fra forskellige nationale registre (CPR-registret, Landspatientregistret, det danske indkomstregistre) med 14 års follow-up på dødelighed; 2) en dansk national kohorte indeholdende alle danske borgere født mellem 1961-1971 med opfølgning mellem 14 og 24 år på kardiovaskulære udfald, diabetes og KOL; 3) The National Survey of Health and Development, en repræsentativ britisk stikprøve født i 1946, som er den længst kørende fødselskohorte i verden med ensomhed som 68-årig som udfald.

Flere hovedresultater er værd at fremhæve. Overordnet indikerer resultaterne, at hvor meget nuværende kvalitet og kvantitet af sociale relationer beskytter mod dårligt helbred afhænger af både oplevelser gennem livet og køn. For det første dokumenteres det, atforekomsten af effektforskelle på tværs af køn afhænger både af den brugte konceptualisering af sociale relationer og det sundhedsudfald, der undersøges. For de fleste aspekter af sociale relationer og de fleste udfald indikerer de fleste studier i denne afhandling begrænset eller ingen evidens for, at sammenhængen mellem sociale relationsmål og sundhed afhænger af

køn. Dermed indikerer resultaterne, at køn kan være en modererende faktor for enkelte sundhedsudfald og for enkelt aspekter af sociale relationer og ikke andre. Konceptualiseringen af nuværende social relationer som en ressource for individet til at forbedre og vedligeholde et godt helbred (individuel social kapital) er det eneste brugte social relationskoncept i denne afhandling, der indikerer, at sociale relationer har forskellig sammenhæng med dødelighed alt efter køn. Resultaterne indikerer, at højere grader af tillid og mere deltagelse i sociale netværk har en svagere sammenhæng med dødelighed for kvinder. For mænd er mere deltagelse i sociale netværk associeret med en højere risiko for at dø.

For det andet udvider afhandlingen horisonten til et livsforløbsperspektiv. Fra dette perspektiv indikerer resultaterne, at negative sociale oplevelser i forskellige livsfaser er associeret med ensomhed blandt britiske 68-årige. Negative sociale oplevelser i barndommen er muligvis en sensitiv periode, og effekter af negative sociale oplevelser ser ud til at akkumulere igennem livet. Ét af hovedresultaterne i Studie IV er, at effekten af mangelfuld social kontakt på ensomhed afhænger af sociale oplevelser tidligere i livet. Yderligere indikerer resultaterne, at en høj nuværende kvalitet af sociale relationer kan mindske den negative effekt af tidligere negative sociale oplevelser på ensomhed.

For det tredje dokumenteres det, at den socioøkonomiske position i barndommen - som en markør for det sociale miljø - er associeret med både kardiovaskulære udfald, diabetes og KOL som voksen. Resultaterne indikerer, at de borgere, der er opvokset i en lavere socioøkonomisk position har en højere risiko for at udvikle de tre ovennævnte livsstilssygdomme. Sammenhængen er stærkest for KOL og er ens for mænd og kvinder. Denne sammenhæng ser ud til at mindskes over tid.

For det fjerde findes det, at sammenhængen mellem socioøkonomisk position i barndommen og risikoen for at udvikle de tre livsstilsrelaterede sygdomme er delvist medieret af den opnåede socioøkonomiske position som voksen. Den største medierende effekt er observeret for KOL med cirka 67% af sammenhængen medieret af den socioøkonomiske position som voksen. Den laveste grad af mediering er observeret for kardiovaskulære udfald med omkring 50% medieret af den socioøkonomiske position som voksen.

For det femte blev det demonstreret, hvordan teknikker fra maskinlæring kan blive brugt inden for sundhedsforskning. Ét er af hovedresultaterne er, at ud af 42 korrelater af ensomhed for 68-årige er både kvaliteten og kvantiteten af sociale relationer blandt de vigtigste prædiktorer af ensomhed.

Ét mål eller én konceptualisering af sociale relationer målt på ét tidspunkt er ikke nok til at give et dækkende billede af vigtigheden af sociale relationer for helbred. Forskellige aspekter af sociale relationer tapper ind i de samme underliggende dimensioner af sociale relationer. Alle disse aspekter af sociale relationer interagerer er med hinanden i at give et overblik over vigtigheden af sociale relationer for helbred. I denne afhandling er det yderligere demonstreret, hvordan dataanalytiske fremskridt i andre områder kan fremme viden inden for social epidemiologi. F.eks. ved at give muligheden for at vurdere den relative vigtighed af mange potentielle prædiktorer af sundhedsudfald. Dette kan bl.a. hjælpe udviklingen og målretningen af sundhedsinterventioner.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

"No man is an island, entire of itself; every man is a piece of the continent, a part of the main."

–John Donne, Devotions upon Emergent Occasions[1]

Human beings have heavily relied on each other since the beginning of our species. Like other mammals in the animal kingdom, our very survival has been and still is, dependent on our inclusion into a social group for protection, sustenance, comfort and development. We shape and are in turn shaped by our social relations, with each social interaction subtly becoming part of our personal sense of self. Over the life course, these millions of social interactions - with family, friends, lovers, teachers, acquaintances and even strangers shape not just who we become but our health in turn.

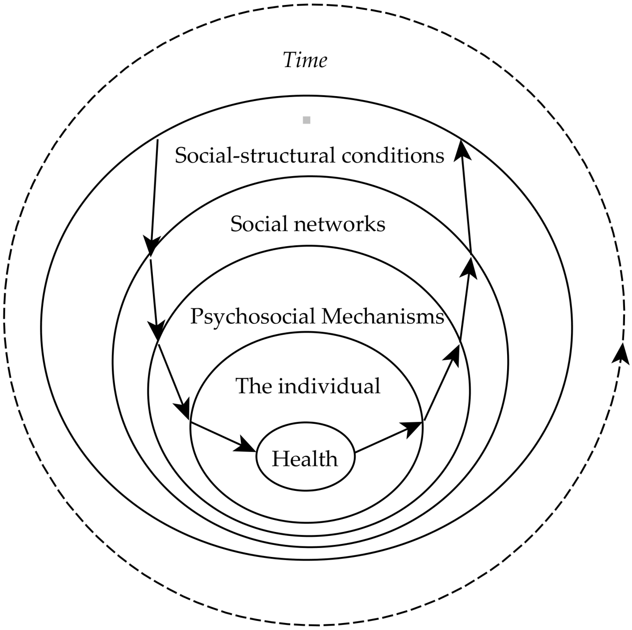
[1]Donne 1923, p. 8

While epidemiology is a scientific discipline that investigates the distribution and determinants of population health, social epidemiology has a special focus on the *social* determinants of health (Miller et al. 2009). Similar to John Donne's statement about the interconnectedness between people, Mervyn Susser noted the interconnectedness between health and people (Susser 1973). This statement stands in stark contrast to the traditional biomedical approach to health, which assumes ill health to be fully accounted for by deviations from the norm of measurable biologic variables (Engel 1977). Instead, social epidemiology recognises that complex social, contextual and political factors all play a part in the individual's propensity to live a fulfilling and healthy life (Krieger 2001; Berkman, Kawachi, and Glymour 2014). Thus, both good health and its absence are not purely biologically determined but socially constructed and distributed. These social circumstances include both proximal and distal social factors in the disease chain across the life course (Berkman, Glass, et al. 2000). Link and Phelan describe social factors as fundamental causes of disease "because they embody access to important resources, affect multiple disease outcomes through multiple mechanisms, and consequently maintain an association with disease even when intervening mechanisms change." (Link and Phelan 1995, p. 80).

A key social epidemiological argument is that not all humans are created equal (Miller et al. 2009, p. 17). Instead, humans are shaped by the social circumstances in which they reside. Across geographical places and time, the social environment shapes and is in turn shaped by the individuals who reside in it, shape and are shaped by the next generation (illustrated in Figure 1.1). This interaction between the individual and the proximal and distal factors in the social environment are often so subtle that it goes unnoticed. One such health-related example is the changing norms in lifestyle behaviours such as drinking, smoking and exercise (Lindbladh et al. 1997; Schooling 2001; Wakefield, B. Loken, and Hornik 2010).

The changing flow and influence of society on the individual is beyond the scope of this dissertation. However, I use this example because it illustrates how health research and social epidemiology are cross-disciplinary fields of research.

Figure 1.1: Extended ecological model of reciprocal influences on individual health across time



Own representation. Adapted from Berkman et al. (2000) and McLeroy et al. (1988) to incorporate a time dimension and the reciprocal nature of the upstream and downstream levels

Important social factors shaping the distribution of adverse health across strata are socioeconomic position, gender, ethnicity and social relationships as well as the broader societal and political context in which an individual lives (Berkman, Glass, et al. 2000). Social relationships are one of the social factors that have emerged within social epidemiology as being crucial for understanding differences in health and longevity (Holt-Lunstad 2018).

Interest in social relationships' impact on health from other disciplinary fields such as sociology, anthropology and psychology emerged long before the rise of epidemiological interest. These theories has helped to lay the groundwork for social epidemiological understanding (Lund 2016). In the late 1890's, the sociologist Durkheim sought to demonstrate how society influences even the seemingly most individualistic act that suicide is. Based on empirical analyses, he theorised that suicide is related to the degree of social integration in society and to a much lesser extent an individual's life circumstances and events. While Durkheim's empirical investigations have been the object of notable criticism, his work has been and still is an important inspiration for social epidemiologist seeking to gain an understanding of the influence of social relations on health (Kushner and Sterk 2005). From psychiatry, John Bowlby's (1979) attachment theory was and still is crucial for understanding how early life attachments continues to exert influence throughout life for health and well-being.

The association between social relationships and different states of health started to gain interest within epidemiology in the 1970's with the studies from Cobb (1976) and Cassel (1976) (House, Landis, and Umberson 1988). Since then, the association between measures of social relationships and different health outcomes has been investigated in a vast number of studies within epidemiology (Holt-Lunstad 2018; House, Landis, and Umberson 1988). This research has amassed compelling evidence for the association between social relationships and both mortality (Steptoe et al. 2013; Holt-Lunstad, T. B. Smith, and Layton 2010; House, Landis, and Umberson 1988), cause-specific diseases and mental well-being (Thoits 2011).

According to Berkman et al. (2014), social relationships impact health through three different pathways: Health behaviours, psychological pathways and physiological pathways. Health behaviours include lifestyle-related behaviours such as smoking, drinking, exercise and diet as well as the degree to which an individual seeks help when needed and adheres to treatments. The psychological pathways include a number of affective states and learnt responses to life events such as the degree of self-efficacy, emotional regulation and self-esteem as well as the degree of depression and coping skills. Lastly, some physiologic responses - primarily associated with bodily stress functions - have been shown to be another way which social relationships influence health (see Figure 1.2 on page 12). This, social relationships can initiate the onset of ill health and in this way serve as a direct cause of both mortality and morbidity through more proximate factors to health (Umberson and Montez 2010).

The question as to how, why and to which extent social relationships influence physical and mental health are thus essential questions in population health. However, despite recognition of the importance of social relationships for health and a longstanding and continued rising interest for the subject (Valtorta et al. 2016), much is still unknown about the complicated interplay between social relationships and health (Thoits 2011; Lund 2016).

Several researchers have highlighted areas for further exploration. Umberson (2010), Berkman & Krishna (2014), Urchino (1996) and Kuh et al. (2003) have emphasised the importance of incorporating a life course perspective on health. The emergence of life course epidemiology has provided a lens through which one can examine how social relationships throughout life influence physical and mental health later in life (Kuh, Ben-Shlomo, et al. 2003). In turn, this perspective may identify opportunities for interventions and enable a better understanding of the social aetiology of ill health. However, studies focusing on social relationships and health in a life course perspective are still few (P. Thomas 2011).

One of the issues of life course studies of social relationships and health is that measurements of social relationships across a lifespan are extremely rare (Lund 2016). The longest running birth cohort in the world (The National Survey of Health and Development) and the Danish Registries are excellent data sources for such investigations. These two datasets enable investigations into different aspects of both health and social relationships (see section 1.3.1) to study potential accumulating or mediating effects and sensitive periods.

Other methodological challenges in measuring and analysing the association between social relationships and health could be furthered by incorporating some of the statistical innovations that have occurred in other fields. Investigating mediation effects is an integral part of assessing the possible pathways through which social relationships influence health (Thoits 2011). However, it is only recently that statistical models have been developed that can validly assess such potential effects in time-to-event data (Lange and Hansen 2011; Lange, Vansteelandt, and Bekaert 2012; VanderWeele 2015). Using such an approach within a life course framework may further the understanding of mediating effects throughout life.

Also, due to the vast number of studies investigating the association between measures of social relationship and health, a high number of predictors and correlates have been identified. Recursive partitioning

- a machine learning technique - could assist in furthering the knowledge on the relative importance of these. This technique could also assist in identifying high-risk groups for specific health outcomes (Strobl, Boulesteix, Zeileis, et al. 2007; Zhang and B. Singer 2013). These two aspects may be used in improving the design and targeting of interventions. This has been highlighted by Berkman (2014) as an important next step in improving population health.

Lastly, several scholars have highlighted the increased need to investigate potential interactions between social factors not only to identify especially at-risk groups but also to identify potential alleviating or buffering factors (e.g. Uchino, J. Cacioppo, and Kiecolt-Glaser 1996; Thoits 2011). Krieger (2003) and Jahn et al. (2017) have emphasized the importance of investigating gender differences within epidemiology. Gender differences in health can occur both due to natural sexual distinctions in biology and due to membership in different social groups. Thus, investigating the buffering effect of gender on the associations between social relationships and health may have its merits in terms of understanding potential differences in how social relationships influence health depending on gender.

In the remainder of this chapter, I first introduce the definition of health and present the different health outcomes that I have investigated in this dissertation (section 1.1). Section 1.2 presents a brief overview of five different aspects of social relationships followed by an argumentation of how these different but related aspects may be seen as tapping into different features of social relationships. In this section, I also briefly present empirical and theoretical argumentation for why and how these social relationship aspects can be thought to influence health. In section 1.3, I argue for how a life course approach and recursive partitioning may help further the knowledge on social relationships and health and briefly present the two approaches. In section 1.4, I summarise the main points of Chapter 1. Lastly, section 1.5 presents the purpose and the specific aims of this dissertation.

1.1 Definition of health

The World Health Organization (WHO) defines health as "... a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity" (World Health Organization 2016). This definition rejects the biomedical mind-body dualism that regards the mind as separate from the body. Instead, a more holistic view of health is embraced. This broad definition regards health as a whole entity and not just the sum of different diseases. In turn, however, this broad definition adds complexity to quantitative studies seeking to gain knowledge that can improve population health. In this dissertation, I investigate four physical health outcomes and one mental health and social well-being outcome.

Markers of physical health A total of 14,199 disease diagnoses are currently listed in the International Classification of Diseases (ICD) - the standard international tool for classification and identification of diseases within epidemiology. The simplest way to summarize the differences in overall health according to some exposure of interest is to use all-cause mortality as the health outcome. All-cause mortality rates have decreased steadily over the past 15 years both globally and locally. However, while the decrease can be observed for both men and women, differences in mortality rates still remain (World Health Organization 2016).

While all-cause mortality is an important marker because it gives an overall view of the how specific groups fare compared to others without having to consider the type of disease, it has some drawbacks. The most important of those is that results based on such a measure will make it hard to pinpoint and target interventions to improve population health. Looking at the leading projected causes of death in 2030 (table 1.1) - cardiovascular disease (CVD), diabetes and chronic obstructive pulmonary disease (COPD) are all in the ten leading causes of the global burden of deaths (Mathers and Loncar 2006). Thus, understanding the social aetiology of these three lifestyle-related diseases continues to be an important point of research in social epidemiology.

Markers of mental health As evident in the definition of health by the WHO, good health is more complex than simply the absence of physical

Table 1.1: Projected 10 leading causes of death in 2030 by the WHO

Rank	**GHE code**	**Cause**	**Deaths (000s)**	**% deaths**	**Deaths per 100,000 population**
1	113	Ischaemic heart disease	9245	13,2	112
2	114	Stroke	8578	12,2	104
3	118	Chronic obstructive pulmonary disease	4568	6,5	55
4	39	Lower respiratory infections	3535	5,0	43
5	80	Diabetes mellitus	2464	3,5	30
6	68	Trachea, bronchus, lung cancers	2413	3,4	29
7	153	Road injury	1854	2,6	22
8	10	HIV/AIDS	1793	2,6	22
9	11	Diarrhoeal diseases	1617	2,3	20
10	112	Hypertensive heart disease	1457	2,1	18

Based on data from the WHO (2016)

disease and infirmity. Good mental health and social well-being are an important aspect of living a healthy and fulfilling life. In this dissertation, I have chosen to focus on loneliness in later adulthood as a representation of this aspect of health. The reasons for looking at loneliness as a health outcome are many.

"Imagine a condition that makes a person irritable, depressed, and self-centred, and is associated with a 26% increase in the risk of premature mortality." (J. Cacioppo and S. Cacioppo 2018). Loneliness has recently been highlighted as an emerging public health problem due to the consistent evidence linking this affective state to several mental and physical health outcomes (J. Cacioppo and S. Cacioppo 2018). Cacioppo and Patrick even go as far as stating that the changes created by loneliness "…are compounded in ways that may be hastening millions of people to an early grave" (2008, p. 5). For older adults, studies show that loneliness is associated with being less physically active, higher rates of alcohol abuse, a greater risk of obesity and sleep deprivation as well as worse cognitive functioning and several adverse physical outcomes (Hawkley 2015; J. Cacioppo and Patrick 2008; Steptoe et al. 2013).

In itself, loneliness is debilitating affective state pointing to a lack of social well-being. Loneliness does not equal lack of social contact in terms of quantity – often referred to as objective social isolation. Instead, loneliness is a subjective feeling arising when the individual perceives that the social relations in which they are embedded does not fulfil their needs. (Hawkley 2015; Hawkley and J. Cacioppo 2010; J. Cacioppo and Patrick 2008; De Jong Gierveld 1987). Consequently, it seems that we cannot equate a low score on objective measures of isolation with a high degree of loneliness. According to de Jong Gierveld (1987) the reason for this lies in the individual's subjective evaluation that mediates and moderates the relationship between objective measures of social relationships and loneliness. Thus, one may lead a life being socially isolated but not feel lonely, while another may feel lonely even while being socially embedded (Hawkley and J. Cacioppo 2010). Thus, investigating the interplay between social relationships and loneliness as a health outcome has its merits in terms of a better understanding the social antecedents of why it arises and how it may be alleviated.

1.2 The Importance of Social Relationships for Mental and Physical Health

In order to adequately assess the importance of social relationships for health, measures are needed that represent the multifaceted and complicated phenomenon that social relationships are (Lund 2016). A great number of different conceptualizations and theoretical approaches to social relationships has emerged across scientific disciplines (Antonucci and Akiyama 1987; Due et al. 1999; Berkman, Glass, et al. 2000; Valtorta et al. 2016) each contributing to an understanding of how social relationships work to influence health. Social relationships are not adequately represented by looking at social relationship aspects in isolation, e.g., how many people we see over a given time period, how often, the quality of these relationships, how we perceive them, the experiences - good and bad - that being in contact with other people will bring, the resources in terms of money, power, prestige or the information we gain access to by being in a given social relation. Instead, all these specific aspects work in tandem across the life course and together shape physical and mental health.

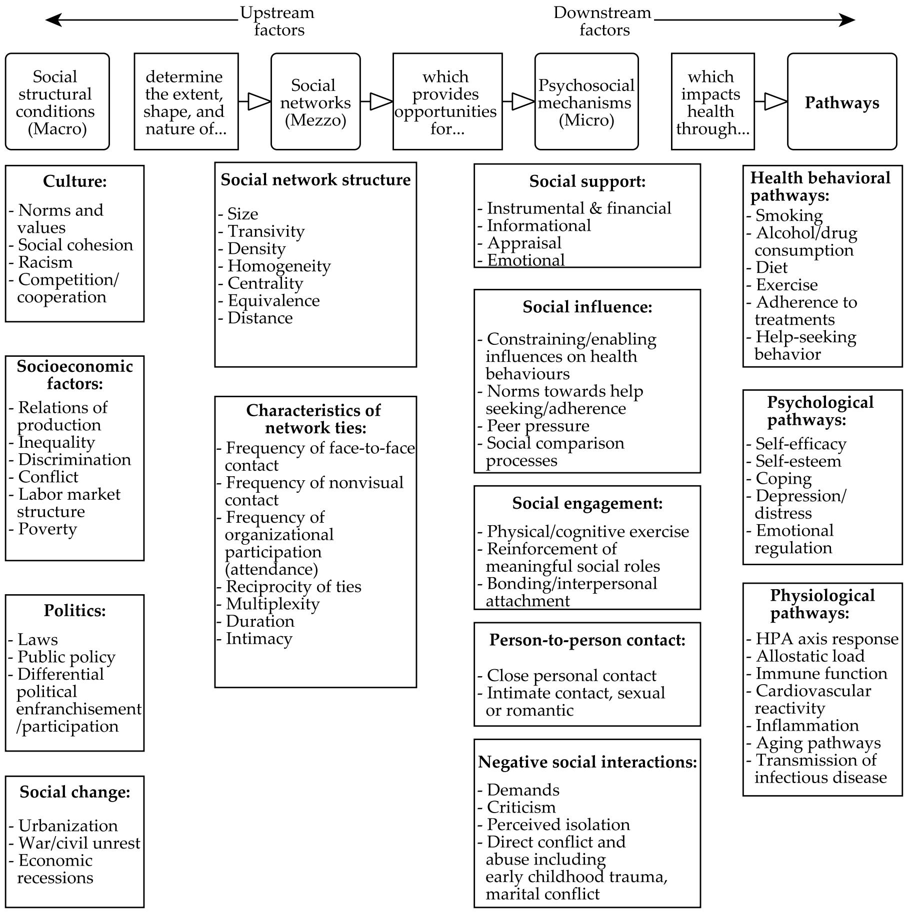
Within social epidemiology a common distinction is made between structural and functional aspects of social relationships regardless of the social relationship concept used (Lund 2016). Structure is related to the quantity and covers duration, frequency, size of social relationships while function taps into more quality related aspects such as social support, relational strains and social integration (Due et al. 1999).

A number of renowned theorists within social epidemiology, sociology and related disciplines have recognised the need for conceptual frameworks that can guide empirical investigations and provide an overview of the many different aspects of social relationships. The most prominent of those is perhaps the conceptual framework by Berkman illustrated in Figure 1.2. This framework is an updated version of the framework presented in Berkman et al. (2000) to also include the downside of social relationships at the micro level (Berkman and Krishna 2014).

This framework illustrates how mezzo and micro social relationship levels are embedded within higher (macro) structural conditions such as culture, politics and socioeconomic factors. These upstream factors condition the opportunities and nature of the factors located downstream from them. According to this framework, social relationship structure and characteristics such as size, the frequency of contact, reciprocity and organizational participation are in part determined by the larger macrosocial context in which they reside. In turn, these structural social networks conditions opportunities for social support, influence, engagement, personal contact, access to different types of health-promoting resources but also opportunities for negative social interactions. These processes then influence behavioural, psychological and physiological pathways to health (Berkman and Krishna 2014).

This model does not incorporate the change and influence of social relationship over time nor does it capture the reciprocal influences these factors have on each other. However, it does provide a theoretical underpinning for understanding the impact of social relationships on health. Focusing solely on social relationships, Due et al. (1999) created a conceptualization distinguishing between structural and functional aspects specifically. In contrast to Berkman's model, the authors disregarded the potentially different proximity to health of the different social relationship factors. Instead, they chose to focus on whether the social relationship factor in question taps into a structural or functional domain. Wang et al. (2017) have created a conceptualization related to social isolation or loneliness. Here, they have argued for how different commonly measured aspects of social relationships tap into different or overlapping underlying features of social relationships. In this way, they have argued that while many social relationship measures are related but distinct aspects, they overlap in the underlying social relationship features that they tap into.

Figure 1.2: Conceptual framework of social relations embedded within social structure showing upstream/downstream social determinants of health



Replicated from Berkman & Krishna (2014, p. 242)

1.2.1 Aspects of social relationships

In this dissertation, the analyses are based on the overall conceptual framework by Berkman (2014) with contributions from the conceptual models by Wang et al. (2017) and Due et al. (1999). I focus primarily on three aspects of social relationships: 1) social relationship adversities (including change over the life course), 2) individual social capital and 3) socioeconomic position in early-life (including change over the life course)[2]. Furthermore, I investigate potential buffering effects on/of 4) the perceived quality of social relationships and 5) frequency of social contact. Lastly, I investigate the relative importance of a broader range of measures that tap into both structural and functional aspects of social relationships (Study III). However, for simplicity, I have only included these five social relationship aspects in this overview.

Valtorta et al. (2016) have commented on the difficulty in comparing results from different concepts and operationalisations. Thus, based on the frameworks by Wang et al. (2017) and Due et al. (1999), I have created an overview of the social relationship concepts used in this dissertation including an indication of the underlying social relationship feature/s that they tap into. The overview can be seen in Table 1.2. The overview provides an indication of whether the social relationship exposures tap into network features relating to either quantity, structure or quality, appraisal of relationships relating to either emotional or resources and lastly experiences relating to either direct or indirect. The overview further provides information on, in which study the aspect in question were investigated (Studies I-IV). In the following paragraphs, I lay out the conceptual and empirical background for each of the five aspects used and argue for which underlying social relationship features that they tap into as well as where in the upstream/downstream chain they reside.

[2] see section 1.2.1 for argumentation on how early-life socioeconomic position can be thought of as a proxy for social relationships in early-life.

Table 1.2: Overview of aspects of social relationships used in this dissertation and the social relationship feature/s they may tap into

Aspects	**Social relationship features**							**Study**
	Network			Appraisal of relationships		Experiences		
	Quantity	Structure	Quality	Emotional	Resources	Direct	Indirect	
Individual social capital				X	X			I
Early-life socioeconomic position		X					X	II
Perceived quality of relationship			X	X				III,IV
Social isolation	X		X	X	X			III,IV
Social relationship adversities			X	X		X		IV

Own representation

Individual social capital Social capital encapsulates different aspects of social relations as well as tying social relations together with a more economic perspective. Social capital may be thought of as some sort of nexus or tie between sociology and economics. Based on the writings on social capital by Bourdieu, social capital is a means for the individual to enhance its position in society relative to other individuals (Bourdieu and Wacquant 1992, p. 119) and can be measured as either a group or an individual attribute (Portes 1998; Diez Roux 2008). Despite disagreement about the particularities and appropriate measurement of the level of social capital, it is generally agreed that social capital stems from the notion that social relationships have a meaningful value for both individuals and communities. As such social capital can be viewed as a form of capital embedded in social relationships that influences the opportunities and limitations an individual will have throughout life (Portes 1998, Portes 2000). That is, from a health perspective, some individuals have access to more information and resources via their social capital. In turn, this enhances their health relative to other individuals with lower social capital (Kawachi and Berkman 2014).

In this dissertation, I focus on social capital as an individual attribute. While individual social capital is not directly mentioned in the upstream/downstream framework, this aspect can be argued to tap into both mezzo and micro level components of social relationships depending on which dimensions of social capital are measured. For example, structural components of individual social capital such as social contact and civic participation (Hyyppä et al. 2007; Macinko and Starfield 2001) fits into the Mezzo-level representing both network characteristics and structure. In contrast, expectations of reciprocity and trust (Abbott and Freeth 2008; Hyyppä et al. 2007; Macinko and Starfield 2001) may represent more psychosocial (micro) mechanisms and appraisals of social relationships.

Early-life socioeconomic position Socioeconomic position is perhaps the most studied social factor within social epidemiology. Overwhelming evidence ties childhood or early-life and adulthood socioeconomic position to both mortality and morbidity (Claussen, Davey Smith, and Thelle 2003; Power, Hyppönen, and G. D. Smith 2005; Stringhini et al. 2013; Becher et al. 2016; Psaltopoulou et al. 2017). While early-life socioeconomic position does not directly indicate some of the quantity or quality of social relationship it may serve as a proxy for the type of socialisation and social stress within the social environment.

The socioeconomic position imposes structural conditions on the social environment in which we live (Wadsworth and Butterworth 2006). Thus, the family environment in which we grow up shape the way we live, socialize and our access to monetary, social and cultural resources (Adler and Kwon 2002). Bourdieu (1986) introduced the concept of habitus to capture the way in which individuals internalize ways of being and acting in the world as dispositions to act. Within epidemiology, habitus shares similarities with the concept of embodiment. Embodiment describes how extrinsic social factors at different life stages become embodied into an individual's body (Krieger 2001). The embodiment of behaviours and cognitive function and dysfunction happens through habituation and learning (Halfon et al. 2002). Cockerham (2005) used these in the concept of early socialization and learning to conceptualize how class circumstances create dispositions to act which in turn predict health behaviours such as alcohol use, smoking, diet and exercise. Health behaviours that are strongly correlated with lifestyle-related diseases. Lawrence (2017) investigated the source of education's strong association with health behaviours and found that when seeking to reduce disparities it might be worthwhile to focus earlier in the life course – starting during or before adolescence in terms of interventions. In part due to the fact that many of the dispositions to behave and the utilization of resources that come from being in a particular socioeconomic position is formed earlier on in life. Similarly, according to Wadsworth & Butterworth (2006, p. 36) the likelihood of adverse behaviours and mental health outcomes related to negative experiences in childhood is greatly increased by poor socioeconomic circumstances in early-life. Thus, socioeconomic position in early-life can be viewed as the structural framework in which a part of the intergenerational transmission of morbidity and mortality occurs through socialization and learned behaviour. In this way, it can be thought of as a proxy of the structural quality of social relationships as well as conditioning (indirectly) social experiences (Krieger 2001; P. Bourdieu 1986; Cockerham 2005). However, it is important to note that socioeconomic position in early-life also conditions many other aspects that increase the likelihood of ill health in adulthood that are not related to social relationships.

Social support and social isolation The access and use of social support are crucial for the capacity to live a healthy life (Berkman and Krishna 2014). As an overall concept social support covers all forms of support both perceived and actually available from one's social network. Umberson & Montez (2010, p. 3) defines social support as "emotionally sustaining qualities of relationships (e.g., a sense that one is loved, cared for, and listened to)". This definition covers both emotional, instrumental, appraisal and informational support. Further distinctions are made between support that is either perceived to be available and adequate or actually received (Berkman and Krishna 2014).

One's closest confidante is an important source of most types of support. Thus, the quality of attachment to one's closest confidante might have powerful effects on health either as a buffer against other strains or as a direct stressor (Umberson and Montez 2010). Thus, perceived quality can be seen as a representation of the micro psychosocial mechanisms as a way of tapping into social support. Social isolation refers to an absence of or infrequent social contact and have been found to have consistent associations with morbidity and mortality (Umberson and Montez 2010). The frequency of social contact represents the mezzo level and taps into some of the characteristics of the social network.

Social relationship adversities While the majority of research within social epidemiology focuses on the influence of positive aspects of social relationships on health, increasing attention is paid to negative features. Lund (2016) has argued for the need to investigate negative social experiences and states that the adverse effect of these experiences may even outweigh the effect of positive interactions. Negative social interactions are not synonymous with a distinct lack of positive relations or experiences. Instead, they are described as direct behaviours towards the individual which are perceived as aversive or unwanted (Brooks and Dunkel Schetter 2011). In this dissertation, I have conceptualized negative social interactions as social relationship adversities. Thus, as social experiences that may have long-lasting effects on the individual such as deaths of close friends or family, divorce, loss of social contact with loved ones etc. Such social relationship adversities may have long-lasting detrimental effects throughout life on both day-to-day activities and social interactions (Jonsson et al. 2016, Dohrenwend 2006).

The social relationship adversities used in this dissertation encapsulate a slightly broader scope of experiences than the conceptualization used by Brooks (2011). Using Berkman's framework, social relationship adversities represent the negative social interactions described under the micro psychosocial mechanisms and tap into features of social relationship regarding the quality and emotional appraisal of direct experiences (see Table 1.2).

There is evidence to suggest a so-called dark side of being in contact with other people. Studies find associations with both physiological stress responses tied to a range of adverse health outcomes as well as both adverse physical and mental health (Rikke Lund et al. 2009; Savla et al. 2013; Umberson, Williams, et al. 2014). Thus, more research focusing on the negative aspects that may arise from being in contact with other people might be an important contribution to the knowledge on the impact of social relationships on health. However, such investigations are difficult due to a lack of reliable access to measures of earlier experienced social relationship adversities as I have mentioned previously in this chapter.

1.3 Methodological opportunities for gaining new insight

New state-of-the-art quantitative methods and conceptual frameworks have been developed that may help to dive further into the relationship between social relationships and health outcomes. I previously highlighted methodological areas that may each help improve knowledge on the relationship between social relationships and health; life course and machine learning approaches. Each of these may help in gaining new insight informing both theory and health interventions in different ways. The utility of these approaches is explained in more detail in section 1.3.1 and 1.3.2 with a brief summary below.

A life course approach can help further the understanding how social relationships occur, accumulate and affect each other over the life course (Kuh, Ben-Shlomo, et al. 2003). This approach can help in understanding the social aetiology of specific health outcomes by enabling investigations that 1) situate social relationship quality and quantity in specific life stages, and 2) investigate the extent to which social relationships at different life stages influence health and social relationships later on. In this regard, a counterfactual approach may be used both within a life course framework to 1) assess possible pathways in which social relationships influence health, and 2) decompose the social relationship effects on health into indirect, direct and total effects (VanderWeele 2015). Lastly, recursive partitioning from the field of machine learning enables 1) assessing the relative importance of a large number of factors in relation to a specific outcome while considering all possible interrelations between them, and 2) identifying which combinations of a large number of factors that best identify high-at-risk groups for specific health outcomes in populations (Scott, Jackson, and Bergeman 2011).

1.3.1 Gaining insight using a life course approach

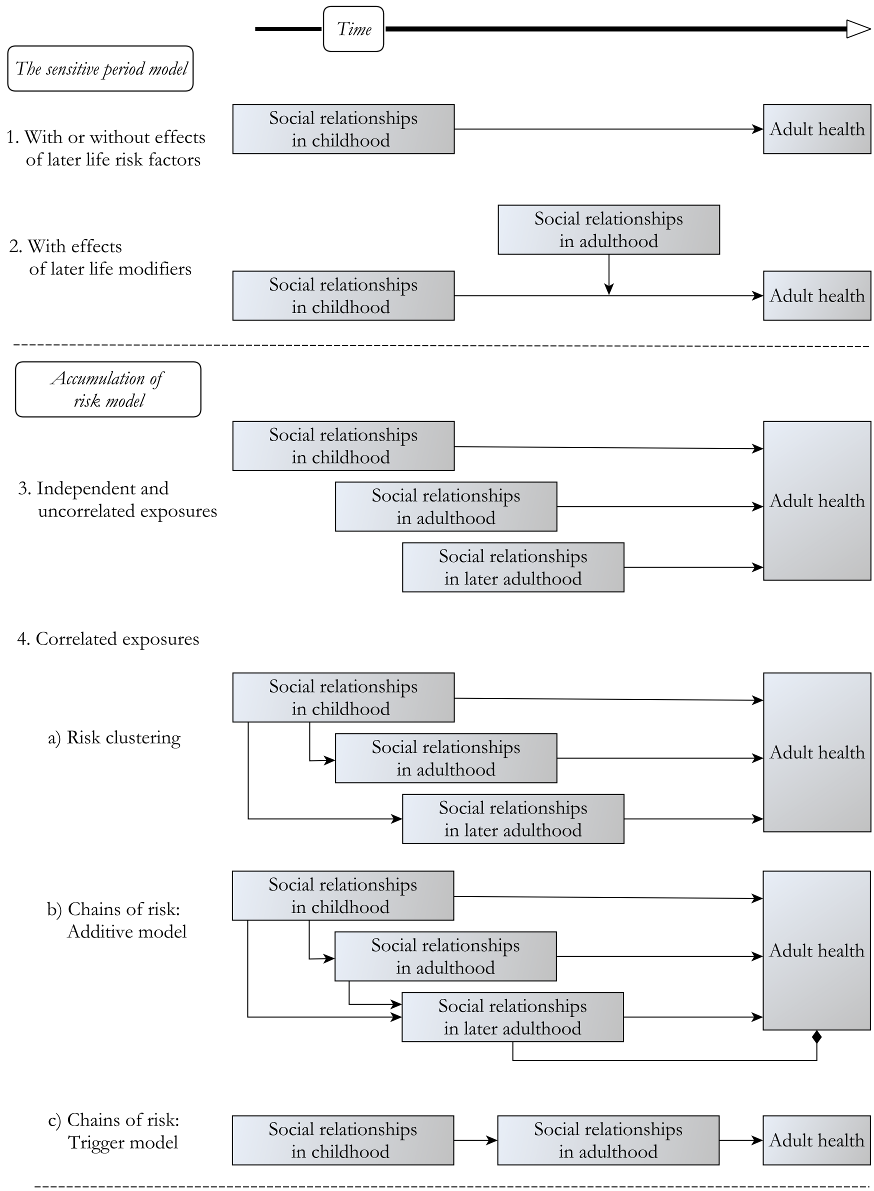
A life course perspective states that variations in health in throughout life are socially patterned and in part shaped by previous experiences (Kuh, Ben-Shlomo, et al. 2003). Giele & Elder (1998) define a life course as "a sequence of socially defined events and roles that the individual enacts over time". In the study of the impact of social relationships on health, a life course approach is the study of long-term effects of social relationship exposures during, gestation, childhood, adolescence, young adulthood and later adult life on health (Kuh and Hardy 2002; Ben-Shlomo and Kuh 2002; Kuh, Ben-Shlomo, et al. 2003). Thus, this perspective could help us understand the role that social relationship determinants have on health from a broader perspective (P. Thomas 2011).

A number of models have been suggested to account for how the timing of exposures influences health (Cable 2014). In Figure 1.3, I have illustrated two life course models in a social relationship framework: The sensitive period model and the accumulation of risk model.

The sensitive life course model with or without effects of risk factors in later life (model 1 in Figure 1.3) and with later effect modifiers (model 2). The first suggests that the social relationships in childhood influence adult health independent of adult social conditions. The second includes that social relationship exposures in childhood interact with exposures in later life which either enhance or decrease the risk of ill health.

The accumulation of risk model assumes that the number, duration and severity of adverse social relationships increase the risk of ill health in later life regardless of the timing. In this model, adverse social relationships may either cause gradual increases in risk without being correlated with each other (model 3, Figure 1.3) or with being correlated (model 4a-c, Figure 1.3) (Mishra, Cooper, and Kuh 2010). The chain of risk model suggests that the influence of childhood social relationships on adult health is through the intergenerational transmission of risk with adult social conditions accounting for the immediate health impact (Mosquera et al. 2017).

Figure 1.3: Life course models with illustrative examples of the impact of social relationship on health



Own representation with inspiration from: Kuh and Hardy 2002; Ben-Shlomo and Kuh 2002; Mishra, Cooper, and Kuh 2010

1.3.2 Gaining insight using machine learning

Most studies use traditional regression techniques to investigate risk factors or correlates of a given outcome – a so-called first generation approach (Cicchetti and Rogosch 1996). In contrast to traditional approaches, recursive portioning are better able to compare the relative importance of many correlates. This technique makes it possible to rank the included correlates according to which are the best identifiers of a given outcomes across many different single analyses. That way, recursive partitioning provides a framework in which combinations of a large number of different correlates can be investigated together in their relation to health outcomes while selecting those best identifying differing levels of an investigated outcomes. Recursive partitioning has previously been used to predict interrelations between life-events and stress (Scott, Jackson, and Bergeman 2011, but is frequently applied in other scientific fields, e.g. genetics, medicine and psychology (e.g. Bureau et al. 2005; Lunetta et al. 2004; Segal, Barbour, and R. M. Grant 2004; Shen et al. 2007). Thus, this approach may also help to gain insight into the more general field of social epidemiology and - in the context of this dissertation - specifically on correlates of loneliness (Victor and Scharf 2005).

1.4 Introduction at a glance

- Both good health and its absence are socially constructed and distributed by proximal and distal social factors in the chain of disease.

- One such key social factor is the social relationships we have throughout the life course making the extent to which and how social relationships influence physical and mental health essential questions within social epidemiology.

- Previous research ties structural and functional aspects of social relationships to both mortality and pretty much all cause-specific physical and mental health outcomes. In this dissertation, I focus on all-cause mortality, loneliness and three lifestyle-related diseases (CVD, COPD and diabetes).

- How to adequately represent and capture the multifaceted phenomenon that social relationship does not have an easy answer. No one measure or conceptualisation can capture all aspects at once. A great number of different conceptualizations and theoretical approaches to social relationships have emerged across scientific disciplines each contributing to understanding how social relationships work to influence health. All these specific aspects work in tandem across the life course.

- In this dissertation, I primarily investigate five related but distinct social relationship concepts that represent both more distal and proximal social relationship factors in the disease chain as well as tapping into different underlying features of social relationships.

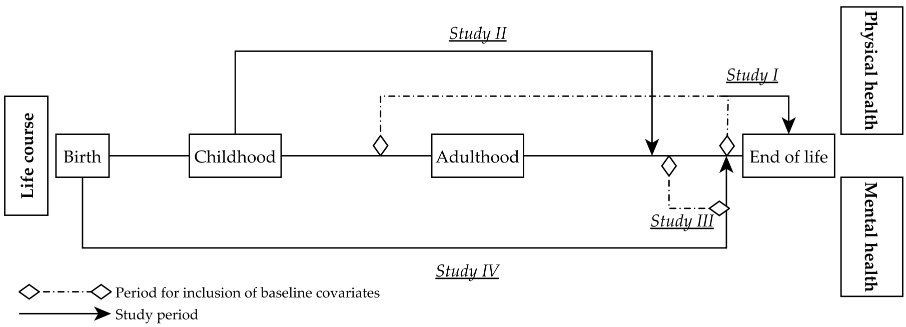
- Life course and machine learning approaches to the quantitative analyses may help to improve the knowledge on the association between social relationships and health.

1.5 Purpose of this dissertation

This dissertation is centred around social relationships and health. The overall aim of this dissertation is to contribute with new knowledge on the importance of social relationships for health. On page 5, I highlighted several areas of research which will guide the sub-aims of this dissertation: 1) Studying the impact of social relationships on health from a life course perspective, including investigating potential buffering or excarberating effects from social relationships earlier on in life, 2) incorporating methodological innovations into the studies of the impact of social relationships on health and 3) potential gender differences in the associations between social relationships and health.

The results are based on four studies (I-IV) for which a life course overview is presented in Figure 1.4.

Figure 1.4: Overview of studies in a life course perspective including indication of whether a study investigated physical or mental health



Own representation

More specifically the studies:

1. Use a life course perspective to investigate influences from
    (a) social relationship adversities at different life stages on loneliness at age 68.
    (b) childhood socioeconomic position of adulthood risk of being diagnosed with lifestyle-related diseases investigating direct, indirect and total influences.
2. Investigate the potential moderating effect of gender on the as-

sociation between structural and functional dimensions of individual social capital and all-cause mortality.

3. Demonstrates how machine learning can be used to advance the knowledge on which aspects of identified correlates are most important for classifying health outcomes.

In Study I, I focus on four specific dimensions of individual social capital and all-cause mortality focusing on potential gender differences. Two sub-studies (Study II and IV) in this dissertation take a life course approach to the relationship between social relations and health (see Figure 1.4). The two studies utilize the life course perspective in order to advance the current literature of how social relationships throughout life influence adult health. The Danish register system via its rich data offers a unique opportunity to overcome many methodological difficulties often encountered in international studies as i.e. common method bias, generalizability, recall bias etc. Thus, the data access to Statistics Denmark is an important aspect of furthering the knowledge of the association between social relations, mortality and mental health. These data are utilized in one of the sub-studies in order to investigate the importance of the socioeconomic position in childhood for adult health and the degree to which this association is mediated by adult socioeconomic position.

While Statistics Denmark has both a range of valid and objective data the access to measures of social relations are limited to public records of cohabitation, marriage, children etc. Thus, in order to investigate the impact of social relationships over the life course in more detail, I perform the other life course study based on data from the National Survey of Health and Development, which is the longest-running survey in the world. This study investigates the association between relations throughout the life course and loneliness at age 68. Study III assess the relative importance of correlates of loneliness in later life using recursive partitioning.

Chapter 2

Materials and methods

"All models are wrong but some are useful"

–George Box[1]

In this chapter, I describe the settings, study designs, materials and methods of the individual studies in the dissertation. Table 2.1 provides an overview of their study design, study populations, outcomes, exposures, covariates as well as the statistical method used in each study. In the following sections, I describe each data source used in the study in greater detail (section 2.1) followed by a section describing the three study populations (section 2.2). An overview of the operationalisation of exposures and covariates follows in section 2.3. Lastly, the analytical strategy is described (section 2.4).

[1] Box 1979

2.1 Presentation of data sources

As stated in Table 2.1 on page 27, the four studies in this dissertation are based on data based solely from survey-based data sources (Study III, IV), register-based data sources (Study II) and a combination of the two (Study I). Below, I briefly outline the Danish National Registers that comprise the register-based data sources (section 2.1.1) and thereafter the two survey-based data sources; North Region Denmark Health Profile (section 2.1.2) and The Medical Research Council National Survey of Health and Development (MRC NSHD) (section 2.1.2). A flowchart illustrating the use of the Danish registries in Study I and II can be seen in Figure 2.1 on page 29.

2.1.1 Register-based data sources

The Danish National Registers

All Danish citizens at birth and people with a Danish residence permit are given a unique civil registration number (CPR) consisting of 10 digits. The CPR is used to identify and collect information on all matters in which an individual has dealings with the public system. Examples include tax, social security issues, place of residence, migration, marriage, divorce, children, hospitalisations, clinical diagnoses, the collection of medicine, place and type of work, education, incarceration and death (M. Schmidt, L. Pedersen, and Sørensen 2014; C. B. Pedersen 2011; Petersson, Baadsgaard, and Thygesen 2011). Statistics Denmark is a public Danish national institution tasked with collecting and processing this information on all Danish citizens, corporations and public institutions (M. Schmidt, L. Pedersen, and Sørensen 2014). This type of data is unique for the Nordic countries and enable access to a wide range of high validity information on an individual level across time for the entire Danish population (C. B. Pedersen 2011). Research institutions and researchers can apply and gain access to anonymous data within one or more of these above mentioned areas after approval from the Danish Health Authorities (Pottegård et al. 2017).

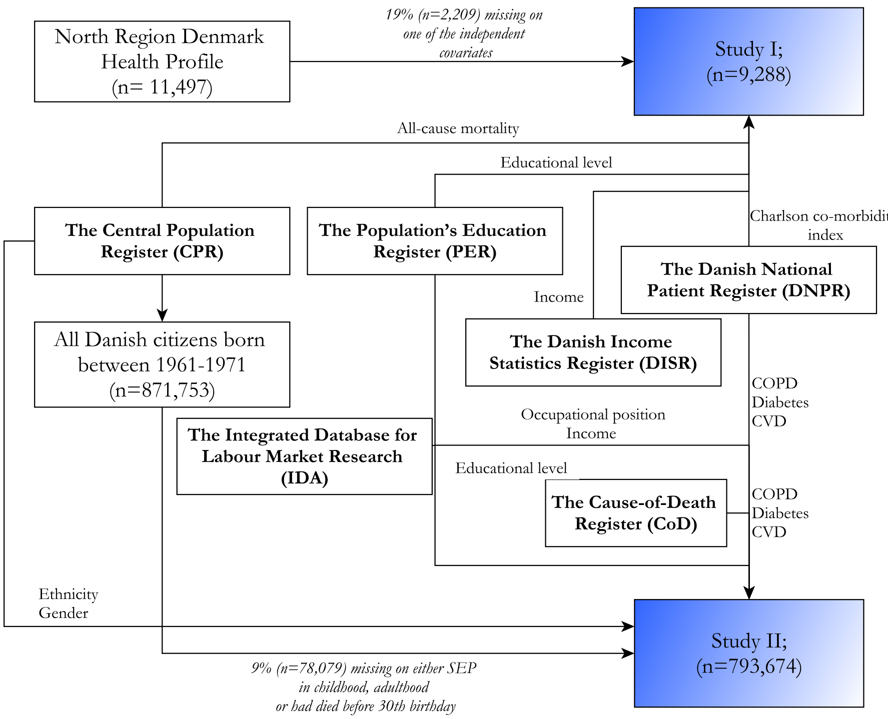
Data from a number of the Danish National Registries were used in Studies I-II. Below is a brief description of the six different registries used in this dissertation. Figure 2.1 shows which covariates have been used from each of the registries described below in Study I and II.

Table 2.1: Overview of the study designs, data sources, study populations, outcomes, exposures, covariates and statistical methods used in the four studies (I-IV). Re=register-based information and S=survey-based information

Study	I	II	III	IV
Study design	Longitudinal study	Longitudinal birth cohort study	Cross-sectional study	Longitudinal birth cohort study
Data source(s)	Survey data and register-based data (Denmark)	Register based data (Denmark)	Survey data (United Kingdom)	Survey data (United Kingdom)
Study population	Participants in the North Denmark Region Survey "How are you?". Men and women, age 16-80 at baseline (2007). n=9,288	All Danish citizens born between 1961-1971. Men and women, age 10-20 at baseline (1981). n=871,753	Participants in the Medical Research Counsel's National Survey of Health and Development (NSHD). Men and women, age 68 at baseline (2014). n=2,453	Participants in the Medical Research Counsel's National Survey of Health and Development (NSHD). Men and women, age 0 at baseline (1946). n=2,453
Outcome(s)	All-cause mortality (Re)	Cardiovascular disease (Re), Pulmonary lung disease (Re), Diabetes (Re)	Loneliness (S)	Loneliness (S)
Exposure(s)	Individual structural and cognitive social capital (S)	Childhood SEP (Re),Adulthood SEP (Re)	42 correlates of loneliness (S)	Number of social relationships adversities in: childhood, mid-adulthood and later adulthood (S)
Covariates	Age, gender, education, income, comorbidity, self-rated health, living arrangement, tobacco use, alcohol consumption, and BMI	Ethnicity, gender, age at baseline	-	Extroversion, neuroticism, childhood social class, serious illness requiring hospitalization for 28 days or more until the age of 20, and gender
Statistical method	Cox-regression including a design weight to correct for sample selection bias	Cox-regression as the natural effect model using an imputation-based approach for nested counterfactuals	Regression Trees and Random Forest	Linear Regression with marginal effects for social relationships measures

The Central Population Register (CPR) includes both historical and current information on dates of birth, death, gender, spouses, place and change of residence (addresses), migration and emigration on every person living in Denmark (C. B. Pedersen 2011). **The Danish National Patient Register (DNPR)** contains information on hospitalisations since 1978 and has an almost complete follow-up. It uses the International Classification Codes (ICD) and links each hospitalisation with the respective information of reason of admission (ICD) to the hospital and date of admission and discharge. This registry is used for hospital statistics, disease monitoring and research. It also contains information on medicine purchases and the type of medicine bought (M. Schmidt, S. A. J. Schmidt, et al. 2015). **The Danish Income Statistics Register (DISR)** holds data on income, taxes, subsidies for all Danish residents. **The Integrated Database for Labour Market Research (IDA)** hold data information on occupational position, levels and annual income (Petersson, Baadsgaard, and Thygesen 2011). **The Population's Education Register (PER)** obtains information on educational degree, type of education and year of completion of the highest educational level (Jensen and Rasmussen 2011). For the Danish studies in this dissertation, we used The International Standard Classification of Education (ISCED) to allow for comparison of educational achievement across countries (OECD 2015). **The Cause-of-Death register (CoD)** hold data on all deaths in the Danish population. Recorded in this registry are a cause-of-death diagnosis using the ICD-codes (Helweg-Larsen 2011).

Figure 2.1: Flowchart illustrating use af Danish registries in Study I and II



Own representation

2.1.2 Survey-based data sources

Besides the Danish registries, two survey-based data sources have been used; The North Region Denmark Health Profile and a British birth cohort entitled the Medical Research Council National Survey of Health and Development. The purpose and overall profile of the two surveys are described below.

The North Region Denmark Health Profile

The North Region Denmark Health Profile is a survey entitled "How are you?". The survey investigates factors relating to health and quality of life, health behaviours including smoking and drinking, social capital and other social relationship measures. The survey used in Study I was conducted in 2007 in Northern Denmark Region. A random selection of 23,490 inhabitants of North Region Denmark was asked to participate in this survey by postal questionnaire (Ejlskov, Mortensen, et al. 2014). These 23,490 inhabitants were drawn randomly from the Civil Registration System stratified by the 11 municipalities in the region. Of the 23,490 inhabitants asked to participate in this health study, 11,497 inhabitants aged 16–80 years responded to either the postal questionnaire or one of two postal reminders that were sent to citizens who had not returned the first postal questionnaire. In total, 48.9% of the inhabitants inquired to participate returned the postal questionnaire (Region Nordjylland and Statens Institut for Folkesundhed 2013).

The Medical Research Council National Survey of Health and Development (MRC NSHD)

The Medical Research Council National Survey of Health and Development (MRC NSHD) is the longest running birth cohort study in the world. The study participants are a representative sample of 5,362 males and females born within marriage during a week in March 1946 in England, Scotland or Wales. It was initially designed to help explain why fertility rates had been falling since the middle of the 19th century and how health services might better the health of both mothers and infants. Over time the scope of the survey broadened and focusses on both physical and mental health, socioeconomic differences and lifetime exposures of the original study participants (Wadsworth, Kuh, et al. 2006). Data have so far been collected on 25 occasions from birth onwards with data being collected approximately every two years in childhood and with main adult collections at ages 26, 36, 43, 53, 60-64 and 68-69 (Kuh, Wong, et al. 2016).

2.2 Presentation of study populations

This section describes the selection of and the basic characteristics of the three study populations used in Studies I-IV. See the attached papers in the appendix for a more detailed description.

2.2.1 The North Region Denmark Health Profile

For Study I, my co-authors and I chose a complete case analysis as the primary analysis resulting in a study sample of 9,288 participants (see Figure 2.1). A range of sensitivity analyses were conducted to investigate whether the removal of participants with missing data on at least one of the covariates influenced the results. I discuss the choice of handling missing data and the potential implications of the different strategies of handling missing data in section 4.1.4 on page 63. The participants were followed from completion of the baseline questionnaire in 2007 until death or study end at the 31st of December 2012. During the five-year follow-up period, 3.5% of the study sample died with men dying at a higher rate compared to women (Ejlskov, Mortensen, et al. 2014).

2.2.2 The 1961-1971 Danish National Cohort

For Study II, the entire Danish population born between 1961-1971 were identified in The Central Population Register (n=871,753). Subsequently, my co-authors and I removed the citizens who had died prior to their 30th birthday from the cohort. Hereafter, all participants who had missing values on at least one of early-life or adult SEP indicators were excluded from the analysis. This resulted in 91% of the original study cohort being eligible for analysis (n=793,674). Following the procedure from Nandi et al. (2012) a study sample for each of the three health outcomes studied (COPD, CVD, diabetes) were then created based on the eligible study cohort. Finally, for each of the three samples, the study participants who had the diagnoses in question before study start at their 30th birthday were excluded. This resulted in a study sample of 792,768 for COPD, 792,675 for CVD and 788,747 for diabetes. During the follow-up period 1.3% developed COPD, 2.6% developed CVD and 4.3% developed diabetes.

2.2.3 The National Survey of Health and Development

As part of the 24th follow-up of the NSHD conducted between 2014 and 2015, the 2942 remaining eligible study members were asked to complete a postal questionnaire (Kuh, Wong, et al. 2016). In total, 83.6% (n=2453) of the study members returned it. This questionnaire includes the 3-item short scale UCLA measure of loneliness (Hughes et al. 2004) that were used as the outcomes in Study III and IV. The questionnaire also includes cross-sectional information on socio-demographic characteristics such as current living situation, marital status, wealth, health indicators, measures on structural and functional social relationships, social and emotional well-being and feelings of mastery. From previous data collections gender, indicators on educational attainment and socio-economic status in mid-life were also included for both Study III and IV. For Study IV, additional information of social relationship adversities throughout the life course was based on previous surveys collected throughout childhood and adulthood with additional information on parental occupation, personality measures and the number of serious illnesses up until the age of 20. The study participants were at the time of the latest collection 68 years old and had a mean loneliness score of 3.8 (range 3-9), with 3 being the lowest loneliness level and 9 being the highest (Ejlskov, Wulff, et al. 2017).

2.3 Operationalisation of outcomes and social relationship exposures

A detailed description of the operationalisation of social relationship exposures can be found in the respective studies in the appendix of this dissertation. Here, I provide a brief description for the social relationship measures used in each study. I also present an overview of the operationalisation of health outcomes for each study.

In Study I, individual social capital was operationalised as four dimensions; two cognitive dimensions and two structural. The choice was based on recommendations from Harpham et al. (2002) and the operationalisation was similar to several previous studies to help cross-study comparison of the findings (i.e, Murayama, Fujiwara, and Kawachi 2012; Nyqvist et al. 2013; Aida et al. 2011). For each of the four dimensions a continuous scale was created by summing the answer to the ordinal-scaled questions outlined in Table 2.2 on page 34. The outcome of this study was all-cause mortality. All-cause mortality was measured with data obtained from Central Population Register with each death of the participants in this study regardless of cause (see Figure 2.1 on page 29 for an illustrative overview of the use of the Danish registries in Study I).

In Study II, early-life socioeconomic position was measured using three different indicators; educational level, job position and income levels (See Table 2.3) (See Table 2.3 on page 35). These three indicators were chosen as they are among the most investigated indicators of socioeconomic position (SEP) (Galobardes, Lynch, and George Davey Smith 2007). Additionally, the indicators were available both in early-life (measured by parental position) and adulthood (measured by own position). The three different lifestyle-related health outcomes; Cardiovascular disease (CVD), COPD and diabetes were identified from the Danish national registries using the International Classification Codes of disease and the ATC-code (for diabetes). From the Cause-of-Death register and the National Patient Registry all citizens were identified with a diagnoses of either CVD (I63, I64, G458, G459, 433,438, I61-64, ICD8: 431-436, I21, I22, ICD10:I21-22, ICD8:410), diabetes (E10, E11, E12, E13, E14, ATC:A10) or COPD (J42, J43, J44, J42, J43, J44).

Study III contained a wide range of covariates. I have given a brief

Table 2.2: Operationalisation of social relationship measures in Study I

Study	Social relationship measure	Items	Categories	Scale
I				
	Individual social capital			
	Interpersonal Trust	How much do you agree with the following statement: "Must people can be trusted"	1:completely disagree, 2:disagree 3: agree, 4:strongly agree, don't know	Continuous
		How much do you agree with the following statement: "Most people try mostly to be fair"	1:completely disagree, 2:disagree 3: agree, 4:strongly agree, don't know	
	Expectations of Reciprocity	How much do you agree with the following statement: "Must people would use you if they got the chance"	1:completely disagree, 2:disagree 3: agree, 4:strongly agree, don't know	Continuous
		How much do you agree with the following statement: "You can't be too careful when dealing with other people"	1:completely disagree, 2:disagree 3: agree, 4:strongly agree, don't know	
	Participation in Social Networks	"How often do you meet with friends that you don't live with?"	1:never, 2;rarely, 3: once or twice a month 4: once or twice a week, 5: daily or almost daily, Don't know	Continuous
		"How often do you meet with family that you don't live with?"	1:never, 2;rarely, 3: once or twice a month 4: once or twice a week, 5: daily or almost daily, Don't know	
	Civic engagement	In your local community "How often do you participate in associations (for example board work, evening school etc.)	1:never, 2;rarely, 3: once or twice a month 4: once or twice a week, 5: daily or almost daily, don't know	Continuous
		"How often do you use the following in your local community: Churches, religious activities, mosques, synagogues	1:never, 2;rarely, 3: once or twice a month 4: once or twice a week, 5: daily or almost daily, don't know	

From Ejlskov, Mortensen, et al. 2014

Table 2.3: Operationalisation of social relationship measures in Study II

Study	Social relationship measure	Items	Categories	Scale
II	**Socioeconomic position in childhood**	Educational level (ISCED)	High: Bachelor or equivalent, master or equivalent or doctoral or equivalent, Middle: Upper secondary and short cycle tertiary, Low: Primary and lower secondary	Ordinal
		Income level (yearly adjusted annual income)	High: Highest tertile, Middle: Middle tertile, Low: Lowest tertile	Ordinal
		Job position	High: non-manual groups: managers and higher salaried employees, Middle: manual groups: skilled and unskilled workers, Low: unemployed and those outside the labour force	Ordinal

overview of the social relationships measures included in Table 2.4 and a detailed description can be found in the appendix (Appendix C). In this study, loneliness was the outcome. Loneliness was measured by the 3-item short UCLA scale (Hughes et al. 2004). In the UCLA scale, participants are asked how often they feel: 1) a lack of companionship, 2) left out and 3) isolated from others. The three items have three response options; 1:hardly ever, 2:some of the time and 3:often. The three items are summed into a composite scale ranging from 3 (the lowest level of loneliness) to 9 (the highest level of loneliness). The validity assessment of this measure in a British setting was undertaken as a part of this study (See Ejlskov, Wulff, et al. 2017) and is available in the appendix (Appendix C).

In Study IV, social relationship adversities were considered in three life stages. First, I went through each of the postal questionnaires collected between 1946 and 2014, in total 24. For each questionnaire I listed the items pertaining to adverse social relationship experiences. Based on the available data on social relationship adversities my co-authors and I then divided the life course into three life stages; childhood (age <18), mid-adulthood (age 36-53) and later adulthood (age 54-64). The included items for each life stage are shown in Table 2.5. Each of the

Table 2.4: Operationalisation of social relationship measures in Study III

Study	Social relationship measure	Items	Categories	Scale
III	**Quality and quantity of social relationships**	Frequency of visits to/by friends or relatives, number friends or relatives seen at least once a month, number of hours of voluntary work per week, number of children (until age 53), how close they live to their nearest child, death of children (until age 37), if they have grandchildren and are regularly visited by them, frequency of participation in social activities ,the identity of the person they felt closest to and the level of emotional support and negative aspects of this relationship	Se appendix C for the specific scoring	Ordinal and continuous

included items were subsequently coded each as having either experienced the social relationship adversity (coded as 1) or not (coded as 0). The exception was the items for being either divorced or widowed where we counted the number experienced. These items ranged from 0 to 2. The items were then summed at each life stage counting the number of social relationship adversities experienced. Due to there being different available items for each life stage, the three counts were standardised to allow for comparison across the three life stages. Similarly to Study III (2017), loneliness was measured by the 3-item short UCLA scale.

Table 2.5: Operationalisation of social relationship measures in Study IV

Study	Social relationship measure	Items	Categories	Scale
IV				
	Social relationship adversities			
	Childhood (age < 18)	Separated from mother for more than three weeks (before age 6)	Yes: 0, No: 1	Continuous summarised scoring
		Difficulties with other children in school (age 9 or 10)	Yes: 0, No: 1	
		Tends to be ignored in school in school (age 13)	Yes: 0, No: 1	
		Unable to make friends (age 13, age 15)	Yes: 0, No: 1	
		Death of mother (before age 18)	Yes: 0, No: 1	
		Death of father (before age 18)	Yes: 0, No: 1	
		Divorce of parents (before age 18)	Yes: 0, No: 1	
	Mid-adulthood (age 36-53)	Number of divorces (age 36-53)		Continuous summarised scoring
		Number of times widowed (age 36-53)		
		Friend/relative died (age 36, age 43, age 53)	Yes: 0, No: 1	
		Lost contact with friend/relative (age 43, age 53)	Yes: 0, No: 1	
		Difficulties with children (age 43, age 53)	Yes: 0, No: 1	
		Serious disagreement friend/relative (age 43)	Yes: 0, No: 1	
		Serious disagreement partner (age 43)	Yes: 0, No: 1	
		Serious disagreements with family or close friends(age 53)	Yes: 0, No: 1	
		Partner had serious accident/illness (age 53)	Yes: 0, No: 1	
	Later adulthood (age 54-64)	Number of divorces(age 54-64)	Count	Continuous summarised scoring
		Number of times widowed (age 54-64)	Count	
		Serious disagreements partner(age 60-64)	Yes: 0, No: 1	
		Serious disagreements with family or close friends (age 60-64)	Yes: 0, No: 1	
		Lost contact with friend/relative (age 60-64)	Yes: 0, No: 1	
		Serious disagreements children(age 60-64)	Yes: 0, No: 1	

2.4 Data analysis

The analyses and data management conducted in this dissertation were performed in either the statistical program R (R Core Team 2017), SAS or Stata with R being the main statistical software program.

2.4.1 Handling of missing observations

Missing data are a potential source of bias in all analyses. I discuss the potential implications of having missing data and compare the different methods used in section 4.1.4 on page 63. In this section, I briefly outline the strategies used to handle missing data in the main analyses for Studies I-IV. In Study I, we used complete case analysis (e.g. listwise deletion). In the main analyses, we further handled the 'don't know' categories of the items used to create the social capital dimensions by directional coding instead of setting them to missing. A battery of sensitivity analyses was carried out to assess whether these two choices affected the results of the analyses (e.g. setting the 'don't know' categories to missing and multiple imputation by chained equations). These analyses resulted in similar estimates compared to the main analyses. I discuss this issue further in the discussion section. In Study II, we also choose to handle missing data on covariates by complete case analysis due to the low missing rate on covariates (Wulff and Ejlskov 2017). In Studies III and IV, we handled missing data by multiple imputation using random forest algorithms (Daniel J Stekhoven 2011; Daniel J Stekhoven and Bühlmann 2012; Daniel J. Stekhoven 2013).

2.4.2 Proportional Hazards (Study I)

The Cox proportional hazard model estimates the hazard rate $h(t)$ for any time t. That is, for an individual i at moment t, the below regression models an individual's hazard $h_i(t)$ compared to a baseline hazard rate and the specified covariates (J. D. Singer and Willett 2003).

$$\frac{h_i(t)}{h_0(t)} = e^{(\beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik})} \tag{2.1}$$

The Cox regression is a semiparametric model with no assumption made about the baseline distribution and is estimated using the partial likelihood. This property is one of the strengths of the Cox-regression and makes the model very flexible. However, the hazard rate $h(t)$ is an average of the rates between an individual $h_i(t)$ and baseline $h_0(t)$ over all time points t. Thus, if the two rates are not proportional over time, the average hazard ratio (HR) is biased to a smaller or greater extent depending on the degree and scale of the lack of proportionally between the two (J. D. Singer and Willett 2003). We found no evidence that the assumption of proportionality was not upheld and calculated HR's with 95% confidence intervals.

We also included several confounding factors which previous studies have shown as possible confounders. These were age, gender, education, income, co-morbidity, self-rated health, living arrangement, tobacco use, alcohol consumption, and self-reported body mass index (BMI). To ensure that standardising the social capital dimensions did not skew the results, we performed an additional analysis that treated both the composite measure of social capital variables and the four specific dimensions as categorical variables with three levels; low, moderate and high. This additional analysis showed similar estimates compared to the main analysis (Ejlskov, Mortensen, et al. 2014).

2.4.3 Proportional Hazards in a counterfactual mediation framework (Study II)

While the concepts of mediation and moderation are nothing new within social epidemiology, using the counterfactual approach in this field may help more validly assess especially the pathways through which an exposure (i.e. social relationships) work (Lange, Vansteelandt, and Bekaert 2012). One of the advantages of using such an approach instead of traditional methods is that the counterfactual approach can be used with all types of statistical models (VanderWeele and Shpitser 2013). To assess to degree mediated by adult SEP, my co-authors and I estimated the total effect (natural direct + indirect effect) of each early-life SEP indicator on COPD, diabetes and CVD, respectively, according to mediation through adult SEP (Lange, Vansteelandt, and Bekaert 2012) The estimation procedure in these analyses follows the procedure outlined in Lange et al. (2017). We used a Cox-regression model as the natural effect model with baseline (time zero) being the day of each of the participants' 30th birthday. We estimated the HR for total, natural direct and natural indirect effects of the different early-life SEP indicatorss on the risk of developing COPD, diabetes or CVD in separate models. For each SEP indicator, we then calculated the percent (%) mediated through adult SEP as the ratio between the natural indirect and the total effect. 95% confidence intervals (95% CI) for the % mediated as well as the total, indirect and direct effect was estimated using 1000 bootstrap samples. To assess potential gender-specific patterns, all analyses were performed on women and men separately. To assess potential differences in the effects across birth year, we conducted a sensitivity analysis stratified by birth year and an analysis with follow up restricted to 14 years for all study participants. These sensitivity analyses showed similar results.

2.4.4 Recursive partitioning (Study III)

Based on multiple sources, I provide a short description of the recursive partitioning techniques applied in Study III below (e.g., Hastie, Tibshirani, and Friedman 2009; James et al. 2014; Strobl, Boulesteix, Kneib, et al. 2008; Strobl, Malley, and Tutz 2009). Through the statistical software environment R (R Core Team 2017), my co-authors and I used the functions available from the R-package tree (Ripley 2016).

Regression trees

In Study III my co-authors and I choose to estimate a regression tree with loneliness as the continuous outcome (Ejlskov, Wulff, et al. 2017). For a regression tree the algorithm uses a recursive binary splitting approach. First, the algorithm starts with a root node containing the entire sample. It then uses the least squares criterion to determine the best cut-off point for all the included predictors. The predictor and its respective cut-off point that minimises the within-group variance of the loneliness outcome produces two groups called child notes. Thus, two groups are generated, each with similar loneliness levels and similar values of the identified correlate. Next, this binary splitting is repeated for each child note at each step identifying the predictor and its respective cut-off that optimally divides the data into two further child notes. For each splitting, the smaller and smaller subsets get more similar both regarding loneliness levels and values of the identified correlates. When the data cannot be divided further, the branches or so-called terminal notes have been reached, and the splitting stops (Hastie, Tibshirani, and Friedman

2009; James et al. 2014; Strobl, Boulesteix, Kneib, et al. 2008; Strobl, Malley, and Tutz 2009).

As the procedure described above is likely to overfit the data, we pruned our trees using cost-complexity pruning. The procedure grows a very large tree and then uses a tuning parameter to control how much the tree should be cut from the bottom-up to result in the lowest test error rate. The tuning parameter that controls this trade-off between the subtree's complexity and its fit to the training data is set through ten-fold cross-validation where the data are randomly split into ten different samples. A tree is grown on nine of the sub-samples while the remaining subsample is used as a test set for calculating the sum of squared error. The procedure is repeated ten times to have each subsample play the part as test and training set resulting in choosing the pruned tree with the lowest cross-validated error.

Improving predictive performance: Random forests

To improve the predictive performance of regression trees we subsequently used the random forests technique. Random forests are an ensemble of single regression trees (see the above section for further explanation). Random forests improve on one of the drawbacks of regression trees where small changes in the data (or even seed) may lead to a different predictor being chosen for the first or subsequent splits (Hastie, Tibshirani, and Friedman 2009; James et al. 2014; Strobl, Boulesteix, Kneib, et al. 2008; Strobl, Malley, and Tutz 2009).

The approach starts by building many regression trees each time varying the sample and the predictors. Thus, for each tree, a fresh bootstrapped sample of the participants and a subsample of the predictors is selected. When a strong predictor (e.g., feeling loved) is left out of a tree, interactions that would have been missed in a normal tree may now be identified. This way of decorrelating the individual trees results in an overall tree-average that is much less variable and thus more reliable than any single tree prediction (James et al. 2014).

We used the R-package randomForest (Liaw and Wiener 2002) and checked that the results are consistent with the cforest procedure in the party package (Hothorn et al. 2006; Strobl, Boulesteix, Zeileis, et al. 2007; Strobl, Boulesteix, Kneib, et al. 2008).

The improved prediction accuracy from random forests comes at the expense of the loss in interpretability as the procedure no longer can be visualised through a single tree. Instead, the results are visualised through variable importance plots. The importance of a given variable is measured through the amount that the residual sum of squares is decreased due to splits over a given predictor, averaged over all the trees. In this case, the selection of important predictors and their relative importance were close to identical to the results when using the procedure suggested by Strobl et al. (2007). In total, the relative ranking of correlates was assessed across 5000 estimated regression trees.

2.4.5 Linear regression (Study IV)

To estimate associations between social relationship adversities at each life stage and loneliness at age 68, we used linear regression through ordinary least squares with loneliness at age 68 as the outcomes. Rubin's rules (1987) were used to combine estimates across the imputed datasets. As part of the analysis, we investigated whether earlier social adversity moderate the association between quantity and quality of current social relationship and loneliness at age 68. Thus, in separate analyses, an interaction term was included between the earlier social adversity and i) frequency of contact with friends, ii) frequency of contact with family outside the household and iii) quality of relationship with the closest confidante, respectively. We then calculated the marginal effects of frequency of contact and social relationship adversities at the three earlier life stages to assess the magnitude of the moderating effect. In this study, we adjusted for the degree of extroversion and neuroticism measured at age 26 (Goldberg et al. 1990), childhood social class measured by father's occupation (Dohrenwend 2006), serious illness requiring hospitalisation for 28 days or more until the age of 20, and gender. Initially, we investigated whether the effects of social relationship adversity depended on gender and found no evidence of a moderating effect.

A range of sensitivity analyses were conducted to assess the robustness of the findings. First, we removed participants with a high degree of missing (>20%) to assess whether the exclusion of these would change the main findings. Second, we assess whether the skewness of the loneliness measure were affecting the results by logistic regression with loneliness classified as scoring 6 or above (Victor and Yang 2012).

Third, we conducted analyses with each of the single social adversity items separately. Each of the sensitivity analyses provided similar conclusions (See Study IV in the appendix for further detail).

2.5 Ethical approval

The Danish Data Protection Agency approved the retrospective register-based study (GEH-2014-014). Ethical approval is not required for a retrospective register-based study in Denmark (Study I and Study II).

With regard to the NSHD, we applied and were granted permission to use the data by University College London. All procedures performed in gathering these data were per the ethical standards of either the institutional and national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. For the data collection occurring at age 68, data collection ethical approval was obtained from the NRES Queen Square REC (14/LO/1073) and Scotland A REC (14/SS/1009)

Chapter 3

Results

"Data!data!data!" he cried impatiently. "I can't make bricks without clay."
– Arthur Conan Doyle[1]

In this chapter, I provide an overview of the main results of the studies in the dissertation (section 3.1). More results, tables and figures as well as a detailed description and discussion of the findings can be found in Studies I-IV (see appendix A-D). In the last part of this chapter, two further analyses are presented focusing on accumulation of risk (section 3.2).

[1]Conan-Doyle 1892

3.1 Results in summary

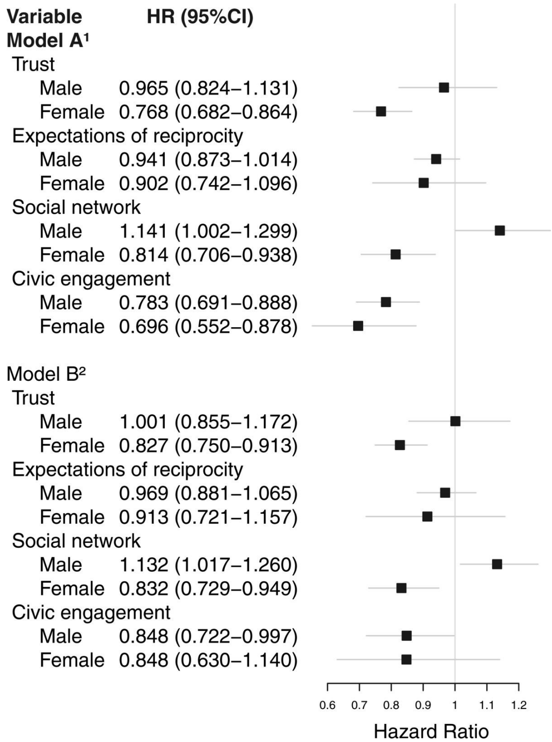
3.1.1 Study I

In Study I, we investigated the association between social capital and all-cause mortality. We investigated both four specific dimensions of social capital (trust, expectations of reciprocity, frequency of social contact, civic engagement) and a composite measure based on the four dimensions. We stratified the analyses by gender based on an initial analysis with statistical significant ($p<0.01$) interaction terms for the composite social capital measure, and the two subdimensions; trust and social network (Table 2 in Appendix A, Study I)

For women, higher levels of the composite measure of individual social capital were associated with lower all-cause mortality. In this study, a one standard deviation increase in individual social capital corresponds to a 43% lower hazard of dying regardless of cause (HR=0.57, 95% CI=0.42-0.83). No such association was found for men (HR=0.95, 95% CI=0.82-1.10) (Figure 2 in Appendix A, Study I).

Analysing the specific dimensions of social capital, the findings suggested that the association with all-cause mortality varied depending on both social capital dimension and gender (Figure 3.1). For women, higher levels of trust and frequency of contact with friends and family were significantly associated with lower all-cause mortality with a one standard deviation increase corresponding to a 17% and 25% decrease in the hazard, respectively (HR = 0.83, 95% CI = 0.75-0.91 and HR = 0.83, 95% CI = 0.73-0.95, respectively). For men, more frequent social contact was associated with a higher hazard of all-cause mortality (HR = 1.13, 95% CI = 1.02-1.26). Civic engagement had a similar effect for both men (HR = 0.85, 95% CI = 0.72-0.99) and women (HR = 0.85, 95% CI = 0.63-1.14) but only reached statistical significance for men Ejlskov, Mortensen, et al. 2014.

Figure 3.1: Associations between social capital dimensions and all-cause mortality (HR(95%CI))



1: adjusted for age and gender. 2: adjusted for age, gender, socioeconomic status (education, living arrangements and income), health status (co-morbidity, self-rated health), and health behaviours (smoking, drinking, BMI). (Ejlskov, Mortensen, et al. 2014)

3.1.2 Study II

In Study II, we investigated the extent of the association between early-life socioeconomic position and COPD, CVD and diabetes, respectively, and the degree to which the association was mediated by the corresponding socioeconomic position in adulthood.

For both men and women, educational position was the socioeconomic marker with the strongest associations with all the three lifestyle-related diseases. The association between early-life educational position and the three outcomes was strongest for COPD (Appendix B, Study II). For COPD, we observed a clear social gradient in the estimated increases in hazard with similar effect sizes for women and men. About two thirds of this association was mediated through adult SEP for women and men. For both diabetes and CVD, the associations was less pronounced. Additionally, the percent of the total effect of early-life SEP mediated through adult SEP for diabetes and CVD was around half for both women and men (Figure 3 in Appendix B, Study II.)

Looking at the number of high positions in early-life, we again observed a clear social gradient with the children who had grown up in families with high positions in both education, income and occupational status having the lowest hazard of developing the three diseases. The extent to which these were mediated by the number of high adult positions all three health outcomes ranged from 32%-43% (Figure 4 in Appendix B, Study II).

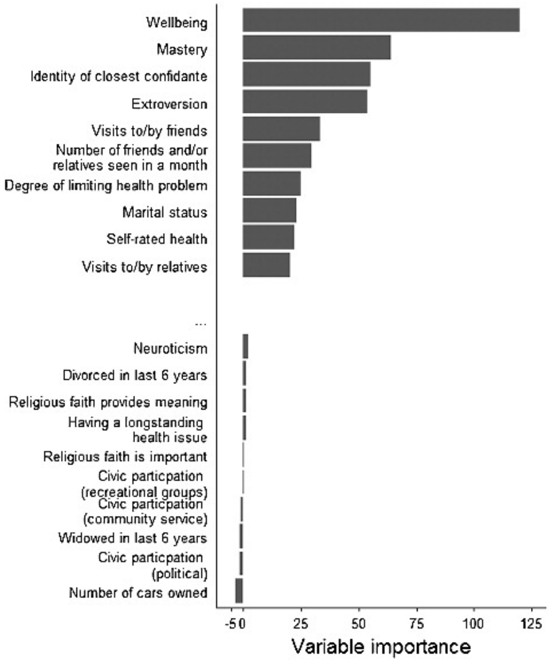
3.1.3 Study III

In Study III, my co-authors and I demonstrated the use of recursive partitioning for exploring and assessing the relative importance of correlates of loneliness in older adults. We included 42 previously identified correlates of loneliness and used regression trees and random forest to assess 1) how combinations of correlates identify subgroups of the population with similar loneliness levels and 2) the relative importance of these correlates in explaining the variation in loneliness in this study sample. Due to only minor differences in the relative importance of correlates between men and women, we choose to perform the analyses on a combined sample including gender as a covariate. Positive mental well-being, personal mastery, identifying the spouse as the closest confidant, being extrovert and informal social contact were the most important correlates of lower loneliness levels. Participation in organised groups and demographic correlates were poor identifiers of loneliness (Figure 3.2). The evidence from the regression tree suggested that loneliness was not higher among those with poor mental well-being if they identified their partner as closest confidante and had frequent social contact. Further, when stratifying on gender, there were only very minor differences between men and women (Ejlskov, Wulff, et al. 2017).

3.1.4 Study IV

In Study IV, we used a life course approach to investigate the association between the number of social relationship adversities experienced in childhood, mid-adulthood and later adulthood and the feeling of loneliness at age 68. Preliminary analyses showed little evidence for a moderating effect of gender and thus, we chose not to stratify the analyses. More recent social relationship adversities were more strongly related to loneliness at age 68 than more distal ones. When considering all confounders and all later social relationship variables included, there was an independent positive association between social relationship adversities at all three life stages and loneliness at age 68 (B=0.04, B=0.07 and B=0.11, in childhood, mid-adulthood and later adulthood respectively). Lastly, looking at the correlation between social relationship adversities at different life stages, social relationship adversity in childhood showed a statistically significant weak association to social relationship adversity in mid-adulthood and was not correlated with

Figure 3.2: Ranking of the 10 most (top) and 10 least (bottom) important correlates across 5000 trees



From Ejlskov, Wulff, et al. 2017

number of adversities in later adulthood. Only, the number of social relationship adversities in mid-adulthood was positively correlated with social relationship adversity in later adulthood (ρ=0.18) (See Table in Appendix D, Study IV.).

3.2 Further analyses

In order to better assess the life course influences of social relationships, I have conducted some further analyses presented below. Section 3.2.1 presents associations between life course accumulation of social relationship adversities and loneliness at age 68, based on data from the NSHD. Section 3.2.2 presents associations between life course accumulation (early-life and adulthood) of three indicators of socioeconomic position and COPD, CVD and diabetes, respectively.

3.2.1 Life course accumulation of social relationship adversities

Looking at life course accumulation (Table 3.1), 1.2% had a higher rate of social relationship adversities in all life stages compared to the remaining 90% of the study population. 4% were in the high group in childhood and mid-adulthood but not later adulthood. 18.5% were in the highest group in childhood but not in mid and later adulthood. 59.3% where in the lower group in all three life stages and 0% of the participants who were in the lower group in childhood were in the high group in mid and later adulthood or low in mid-adulthood but high in later adulthood.

Table 3.1: Life course accumulation of social relationship adversities i childhood, mid-adulthood and later adulthood

Trajectories	n(%)
High High High	20(1.2%)
High High Low	55(4.0%)
High Low High	91(4.9%)
High Low Low	343(18.5%)
Low High Low	280(12.2%)
Low Low Low	1664(59.3%)
Low High High	0(0%)
Low Low High	0(0%)

n is based on the complete case data. Percent(%) is based on all 20 imputed data sets.

Table 3.2 shows the results for life course accumulation of social relationship adversities. All combinations of social relationship adversities have higher loneliness levels compared to having a low degree of social relationship adversities at all life stages. Those participants who experienced a high degree of social relationship adversities at all three life stages had the largest association with higher loneliness levels (β=0.83) followed by those who had high in childhood and mid adulthood and low in later adulthood (β=0.42) and those who had low in childhood and later adulthood and high in mid adulthood (0.24). The remaining two groups did not have a reliable difference in loneliness levels from the reference group.

Table 3.2: Multiple linear regression coefficients (β) for the accumulation of social relationship adversities over the life course predicting loneliness at age 68

Trajectories over the life course*	β(se)	t
Low Low Low	Ref	
High High High	0.83 (0.29)	2.9
High High Low	0.42 (0.17)	2.4
High Low High	0.00(0.14)	-0.0
High Low Low	0.04 (0.08)	0.6
Low High Low	0.26 (0.08)	3.2

**controlled for current social contact with friends and family, quality of relationship with closest confidante, gender, extroversion, neuroticism, childhood social status and number of childhood illnesses*

3.2.2 Early-life and adulthood accumulation of socioeconomic position

Table 3.3 shows the percentage of participants in the different life course accumulation categories of the three indicators of SEP; education, income and job position who had developed either CVD, COPD or diabetes. For all three outcomes a clear gradient is observed for both men and women. Those ending up in a high position regardless of SEP indicator had a lower prevalence of CVD, COPD or diabetes at the end of the follow-up period.

Table 3.3: Descriptive statistics of life course accumulation categories of education, income and job position, respectively

	CVD		COPD		Diabetes	
	Men (n=13932)	Women (n=9031)	Men (n=5770)	Women (n=6064)	Men (n=22950)	Women (n=18389)
Educational position						
High to High	1.5%	1%	0.4%	0.4%	2.1%	2.3%
High to Low	3.8%	2.7%	2.1%	2.3%	5.9%	5.8%
High to Middle	2.3%	1.5%	0.7%	0.8%	3.7%	3.1%
Low to High	1.9%	1.4%	0.4%	0.6%	3%	3%
Low to Low	4.9%	3.9%	2.7%	3.7%	9%	7.9%
Low to Middle	3.3%	2.2%	1.1%	1.1%	5.2%	4.2%
Middle to High	1.8%	1.3%	0.4%	0.5%	2.7%	2.7%
Middle to Low	4.3%	3.4%	2.3%	3%	7.5%	6.7%
Middle to Middle	3%	1.9%	1.1%	1%	4.7%	3.9%
Income position						
High to High	2.1%	1.3%	0.4%	0.5%	3%	2.4%
High to Low	3.3%	2.1%	1.9%	1.9%	5.6%	4.4%
High to Middle	2.7%	1.7%	0.9%	0.9%	4.1%	3.4%
Low to High	2.4%	2%	0.7%	0.6%	4.2%	2.7%
Low to Low	4.8%	3.6%	3.3%	3.7%	8.4%	7.6%
Low to Middle	3.6%	2.2%	1.3%	1.3%	5.8%	4.4%
Middle to High	2.6%	1.8%	0.7%	0.7%	3.9%	3.3%
Middle to Low	4.8%	3.7%	3.4%	3.2%	8.7%	7.6%
Middle to Middle	3.5%	2.2%	1.3%	1.3%	5.8%	4.6%
Job position						
High to High	2.5%	1.4%	0.7%	0.6%	3.7%	2.9%
High to Low	3.2%	2.3%	2%	1.7%	5.7%	4.6%
High to Middle	3.2%	1.9%	1.1%	1%	5.2%	3.7%
Low to High	2.7%	1.6%	0.8%	0.6%	4.5%	2.8%
Low to Low	4.4%	3%	2.8%	2.8%	7.8%	6.1%
Low to Middle	3.8%	2%	1.4%	1.1%	5.9%	3.9%
Middle to High	2.8%	1.5%	0.8%	0.8%	4.4%	3.3%
Middle to Low	4.3%	2.9%	2.7%	2.3%	7.5%	6.3%
Middle to Middle	3.6%	2%	1.3%	1%	6.1%	4.3%

Table 3.4 shows the Cox-regression estimates (HR with 95%CI) of the association between the accumulation categories of the three indic-

ators of SEP in early-life and adulthood and CVD, COPD and diabetes, respectively controlled for ethnicity and birth-year. All accumulation categories of SEP had higher hazards of developing CVD, COPD or diabetes compared to those who both grew up in a high and ended up in a high position regardless of SEP indicator. For all outcomes and SEP indicators, those with a high -> high trajectory had the lowest hazard. Similarly, those who ended up in a high SEP regardless of growing up in a low or middle SEP environment followed in having the lowest hazard ratio. Evidence of gender differences in the strength of the associations differed depending on the SEP marker and indicator used. However, the associations for both genders where in the same direction and were ranked similarly regarding the gradient in associations with the three outcomes.

Table 3.4: Cox-regression estimates (HR with 95%CI) on the association between the life course trajectories of education, income and job position and CVD, COPD and diabetes, respectively on the eligible Danish birth cohort sample (1961-1971) for each health outcome*

	Educational position		**Income position**		**Job position**	
Trajectories	Men HR(95%CI)	Women HR(95%CI)	Men HR(95%CI)	Women HR(95%CI)	Men HR(95%CI)	Women HR(95%CI)
COPD						
High to High	*Ref*	*Ref*	*Ref*	*Ref*	*Ref*	*Ref*
High to Low	3.1 (2.8-3.6)	5.4 (4.2-7)	3.1 (2.8-3.6)	2.6 (2.2-3.1)	5.2 (4.3-6.3)	4 (3.3-4.8)
High to Middle	1.6 (1.4-1.9)	2 (1.6-2.5)	1.6 (1.4-1.9)	1.4 (1.2-1.7)	2.3 (1.9-2.7)	1.9 (1.6-2.3)
Low to High	1.1 (0.9-1.2)	1.2 (0.9-1.6)	1.1 (0.9-1.2)	0.9 (0.7-1.1)	1.5 (1.1-1.9)	1 (0.8-1.4)
Low to Low	4.0 (3.5-4.4)	8.1 (6.7-9.8)	4.0 (3.5-4.4)	4.0 (3.4-4.6)	8.3 (6.9-10.1)	7.4 (6.2-8.8)
Low to Middle	2.0 (1.7-2.2)	2.4 (2-2.9)	2.0 (1.7-2.2)	1.5 (1.2-1.7)	3.3 (2.7-3.9)	2.6 (2.2-3.1)
Middle to High	1.2 (1-1.4)	1.1 (0.9-1.4)	1.2 (1-1.4)	1.1 (0.9-1.4)	1.6 (1.3-2)	1.3 (1-1.7)
Middle to Low	4.3 (3.8-4.8)	7 (5.8-8.5)	4.3 (3.8-4.8)	3.3 (2.8-3.8)	9 (7.5-10.8)	6.5 (5.5-7.8)
Middle to Middle	1.9 (1.7-2.1)	2.3 (1.9-2.7)	1.9 (1.7-2.1)	1.4 (1.2-1.7)	3.3 (2.7-3.9)	2.7 (2.3-3.2)
CVD						
High to High	*Ref*	*Ref*	*Ref*	*Ref*	*Ref*	*Ref*
High to Low	1.3 (1.2-1.4)	2.3 (1.9-2.8)	1.3 (1.2-1.4)	1.6 (1.4-1.7)	1.6 (1.4-1.7)	1.7 (1.5-1.9)
High to Middle	1.2 (1.1-1.3)	1.4 (1.2-1.6)	1.2 (1.1-1.3)	1.3 (1.1-1.4)	1.3 (1.2-1.5)	1.4 (1.3-1.6)
Low to High	1 (0.9-1.1)	1.2 (1-1.4)	1 (0.9-1.1)	1.1 (0.9-1.2)	1 (0.9-1.1)	1.4 (1.2-1.6)
Low to Low	1.6 (1.5-1.7)	3.2 (2.9-3.6)	1.6 (1.5-1.7)	1.8 (1.7-2)	2.1 (1.9-2.3)	2.7 (2.4-3.1)
Low to Middle	1.3 (1.2-1.4)	1.8 (1.6-2.1)	1.3 (1.2-1.4)	1.3 (1.1-1.4)	1.6 (1.5-1.8)	1.7 (1.5-1.9)
Middle to High	1.1 (1-1.2)	1.2 (1.1-1.4)	1.1 (1-1.2)	1 (0.9-1.2)	1.2 (1.1-1.3)	1.3 (1.2-1.6)
Middle to Low	1.6 (1.5-1.7)	2.9 (2.6-3.3)	1.6 (1.5-1.7)	1.9 (1.7-2.1)	2.3 (2.1-2.5)	2.9 (2.6-3.2)
Middle to Middle	1.3 (1.2-1.4)	1.7 (1.5-1.9)	1.3 (1.2-1.4)	1.3 (1.2-1.5)	1.7 (1.6-1.8)	1.8 (1.6-2)
Diabetes						
High to High	*Ref*	*Ref*	*Ref*	*Ref*	*Ref*	*Ref*
High to Low	1.6 (1.5-1.7)	2.5 (2.2-2.9)	1.6 (1.5-1.7)	1.5 (1.4-1.7)	2 (1.8-2.2)	1.9 (1.7-2.1)
High to Middle	1.4 (1.3-1.5)	1.3 (1.2-1.5)	1.4 (1.3-1.5)	1.2 (1.1-1.3)	1.5 (1.4-1.6)	1.5 (1.3-1.6)
Low to High	1.1 (1.1-1.2)	1.2 (1.1-1.3)	1.1 (1.1-1.2)	0.9 (0.8-1)	1.3 (1.1-1.4)	1.1 (0.9-1.2)
Low to Low	2 (1.9-2.1)	3.3 (3-3.6)	2 (1.9-2.1)	2 (1.8-2.1)	2.9 (2.7-3.1)	3.1 (2.8-3.4)
Low to Middle	1.5 (1.4-1.6)	1.7 (1.6-1.9)	1.5 (1.4-1.6)	1.2 (1.1-1.3)	2 (1.9-2.2)	1.8 (1.7-2)
Middle to High	1.2 (1.1-1.2)	1.2 (1.1-1.3)	1.2 (1.1-1.2)	1.1 (1-1.3)	1.3 (1.2-1.4)	1.3 (1.1-1.4)
Middle to Low	2 (1.9-2.1)	2.8 (2.6-3.1)	2 (1.9-2.1)	2 (1.9-2.2)	3.1 (2.9-3.4)	3.2 (2.9-3.5)
Middle to Middle	1.5 (1.5-1.6)	1.6 (1.5-1.7)	1.5 (1.5-1.6)	1.4 (1.3-1.5)	2.1 (2-2.2)	1.9 (1.8-2.1)

** n=788,747 for diabetes, n=792,768 for COPD and n=792,675 for CVD*
all regression estimates are controlled for ethnicity and birth-year

3.2.3 Overview of gender differences

Table 3.5 provides an overall overview of the main results regarding the different social relationship exposures and their respective associations with the investigated health outcomes for men (M) and women (W). For most aspects of social relationships and most outcomes, the studies show limited or no evidence that the association between social relationship measures and health are buffered by gender. Study I showed differing effects for men and women for two of the individual social capital measures. Study II showed some differences depending on gender but the gradient in estimates across categories and direction of associations were the same. Studies III and IV provided no or very little evidence that the associations were different depending on gender.

Table 3.5: Overview af associations (+=positive, -=negative) for men (M) and women (W) for studies I-IV

		Health					Study
		Physical health				Mental health and social wellbeing	
		All-cause mortality	COPD	CVD	Diabetes	Loneliness	
Exposures							
Social relationship adversities							
Childhood						M+,W+	IV
Mid-adulthood						M+,W+	
Later adulthood						M+,W+	
Perceived social support						M+,W+	III
Social isolation						M+,W+	
Individual social capital		W+					I
Cognitive social capital	Trust	W+					
	Expectations of reciprocity						
Structural social capital	Social network	M-,W+					
	Civic engagement	M+					
Socioeconomic environment in early-life			M+,W+	M+,W+	M+,W+		II

M=men,W=women

Chapter 4

Discussion of methods and materials

> *"The key is in remembering that a model is a tool to help us understand the complexities of the universe, and never a substitute for the universe itself."*
>
> – Nate Silver[1]

In the following chapter, I discuss several methodological issues that are important for the interpretation and generalisability of the findings of the studies in this dissertation. I discuss the strengths and weaknesses concerning the study designs, and study populations used,and the issues of missing data (section 4.1). The validity of the findings is discussed in terms of potential attrition, selection bias, information bias, confounding and generalisability. I briefly discuss the implications of the measures of social relationships used in section 4.2. Finally, I discuss and compare implications of the choice of two different statistical approaches - data models vs algorithmic models - in section 4.3.

[1] Silver 2012

4.1 Study design, study populations and the issue of missing

In this dissertation, the study designs were all quantitative. The studies are solely observational studies that are both longitudinal cohort studies and a cross-sectional study (See Table 2.1 on page 27 for a more detailed overview).It can be argued that quantitative studies are ill-suited to capture the multifaceted phenomenon of social relationship and instead calls for in-depth qualitative research. However, while the concept of social relationships is simplified by numerical representations of various dimensions (see section 1.2.1 on page 13), it also affords many advantages over a qualitative approach. First, it enables a more comprehensive and understandable description of the phenomenon in question. By using numerical representations of different dimensions of social relationships, as, i.e. the frequency of social contact, it becomes possible to get an overview of how different dimensions are associated with health. Second, a quantitative approach is well-suited for testing the salience of specific hypotheses in relation to theories as needed to answer the first and third aim of this dissertation. In comparison to a qualitative design, a quantitative study design allows the investigation of social relationship risk factors and the extent to which they influence or are associated with health.

4.1.1 Study designs and issues of causality

Studies I, II and IV are longitudinal since the social relationship exposures are measured before the outcome. Study III is cross-sectional because the social relationship exposures are measured at the same time as the outcome (loneliness) (Rothman and Greenland 1998).

The cross-sectional design in Study III makes it difficult to establish a clear direction of causality. That is, whether it is the included exposures that influence loneliness levels or the other way around. For this reason, we refrained from implying causal directions in the interpretation of the results. Parts of the previous literature has found some of the exposures included in this study to be outcomes of loneliness levels. However, the goal of Study III is not to establish a direction of causality. Instead, it is to demonstrate how machine learning may be used to gain insight into the complicated interplay between a range of exposures. Being able to situate the exposures prior to the participants' loneliness levels would be an added strength of this study. Done this way, it would add to informing better intervention strategies and targeting. Due to the data source being the NSHD, it was possible to include social relationship measures from earlier on in life. My co-authors and I decided against it on the grounds that the only available loneliness measure was measured - at the time - at the latest data collection when the participants were age 68. To date, very few studies have been done that investigate correlates of loneliness in later life using a life course perspective. For this reason, we decided that the argumentation be too weak to include social relationship measures from previous life stages.

In contrast to Study III, the longitudinal framework of Studies I, II and IV strengthen a causal interpretation of the results. However, while we have temporal ordering of the measured exposures and outcome, there may still be issues of temporal dispersion, reverse causality from unmeasured outcome measures at previous life stages (as, e.g. loneliness in Study IV) as well as the ever-persisting problem in observational studies of unmeasured confounding. In summary, while the longitudinal framework of the three studies strengthens a causal interpretation, it should still be done with caution.

4.1.2 Study populations - recall bias, attrition, selection bias and generalisation

To satisfy the first and third research aims, we conducted analyses based on three different study populations that all had longitudinal features. A prospective longitudinal study design is the superior design when wanting to trace changes over time and being able to situate exposures and outcomes at different time points. For this reason, the prospective longitudinal design was an obvious choice for the investigation into the dynamics of social relationships and health over time (Studies II and IV).

Overall, the studies had very different sample sizes. For this reason, the statistical precision of the estimated associations varied considerably across studies (Rothman and Greenland 1998). For example, in Study II, the study population was a Danish nationwide population born between 1961-1971 (n=793,674). Thus, the statistical precision of the estimates in this study is very high. In contrast, Studies I, III and IV were population-based samples consisting of n=9288 and n=2453 participants, respectively. While still large in sample sizes, these populations will undoubtedly have relatively lower precision.

Compared to a retrospective and cross-sectional design, prospective cohort studies are considered more reliable since potential recall biases are much smaller (Mann 2003). For example, in Study IV adverse social relationship experiences were collected from the respective life stages ruling out potential recall biases present in other studies. However, because data were collected over a long period and the questions were not designed specifically to measure social relationship adversities, it is very likely that important adverse social relationship experiences have been missed. Also, it is a possibility that participants who experienced more social relationship adversities in earlier life stages were less likely to respond to the questionnaire regarding loneliness at age 68.

In Study I, no measurement was used that retrospectively delved into previous emotions or experiences. Similarly for Study II, by including the most commonly used SEP indicators from the national registers measured in both early-life and adulthood some of the issues of using retrospectively assessed SEP and potential self-reporting bias were avoided. However, the use of early-life SEP indicators measured at one single time point has been argued to be a weak proxy for the SEP spanning the entire childhood period (Glymour 2007). In contrast to Study II, some self-reporting bias might be present in both the measures used from the NSHD and The North Region Denmark Health Profile due to the sensitive subjects that the participants were inquired about (i.e. loneliness, trust, social contact). Thus, it is a possibility that there is a self-reporting bias that might have resulted in under-reported loneliness levels and that social contact and trust have been over-reported. However, this will only affect the estimated associations if the self-reporting bias is also associated with the other exposures or outcomes measures (Fadnes, Taube, and Tylleskär 2009).

For both Study I and II, data on mortality, lifestyle-related diseases, SEP and most of the confounding variables originated from the Danish national registers. Both the accuracy and completeness of the register-based data are generally considered to be high. The measures and data from the Danish registries are continuously used and updated by the administrative systems in the Danish healthcare system, for reimbursements and other administrative issues which helps to ensure high levels of precision in how well the data is recorded (Thygesen and Ersbøll 2011). Thus, compared to survey-based measures, variables originated from the national register may reduce the risk of differential misclassification.

A general issue with prospective cohort studies is that it takes a long time to collect useful data. The extended collection period makes the data very sensitive to attrition or loss to follow up (Mann 2003). For the NSHD, used as the data source in Studies III and IV, potential attrition bias is a possibility. Analyses at age 60-64 show that the NSHD was broadly comparable to the UK population of the same age on social class and unemployment rate but over-represented owner-occupiers (Stafford et al. 2013). Thus, it may be possible that those from a lower socioeconomic position has different loneliness levels compared to those who did not respond. At age 68, the levels of loneliness in the NSHD are comparable to those from the English Longitudinal Study of Ageing (Ejlskov, Wulff, et al. 2017), but socioeconomic advantage was associated with a higher likelihood of responding and affective symptoms were not (Stafford et al. 2013). For Study I and II loss to follow-up may also constitute a significant risk of bias. However, for both studies, the loss was on a small scale and is thus unlikely to bias the results in any significant way.

The use of a nationwide population cohort in Study II meant that a potential selection bias influencing the findings was not an issue in this study. However, selection bias might be an issue in the three remaining studies (I, III and IV) (Hernán, Sonia Hernández-Díaz, and Robins 2004). The NSHD is based on a representative sample of males and females born in 1946 within marriage during the same week in March of that year in England, Scotland or Wales (n=5,362) (Kuh, Wong, et al. 2016). Children born outside of marriage are not represented in this study. It may be that children from these families growing up in the 1950'ties and 1960'ties experienced more social relationship adversities in childhood and perhaps later. The sample used in Study I - The North Region Denmark Health Profile - was drawn from a less urbanised region in Denmark. Additionally, there were some issues of non-response with a slight over-representation of women and older age groups in this study. Thus, a possible non-response bias cannot be ruled out which may attenuate the association between social capital and all-cause mortality and in that way affect the validity of the results.

For Studies III and IV, we can rule out potential bias resulting from age differences since all participants were born in the same week in March 1946. However, generalising these results to other birth cohorts should be done with caution. Similarly, for Study II, the participants were born between 1961-1971. To ensure that birth cohort effects did not influence the results of this study, we conducted sensitivity analyses stratified by birth year and did not find marked differences between the different cohorts.

4.1.3 Confounding

A significant source of error to consider in epidemiological studies is confounding. Confounding is an issue when some third causal factor is either exaggerating or diluting the relationship between the exposure and outcome of interest (Kish 1959). In the presence of one or more confounders, we risk either attributing too high or too low an effect of, e.g. social capital on mortality. If we can identify and control for confounders accurately, we will to some extent be able to reduce or eliminate confounding bias (VanderWeele and Shpitser 2013).

It is possible to control for confounders in two principal ways: Via design or analysis (Pearl, Robins, and Greenland 1999). A powerful means of design control is randomisation. By randomly assigning individuals to different treatment and control groups, the confounding problem can be eliminated. However, design-based methods are subject to practical limitations that often make them practically infeasible. For example, in Study I, we do not randomly assign people to different levels of social capital and then observe the causal effect on all-cause mortality. Instead, we observe the levels of social capital that the individuals in our sample already have. Thus, to study the connection between social factors and health, an alternative method for approaching the confounding problem is needed.

For the studies in this dissertation, my co-authors and I have relied on analysis-based methods to control for confounding. In Studies I, II and IV, this is primarily done using regression strategies where parametric or semi-parametric constraints are imposed when regressing the outcome in question on the chosen covariates (Rothman and Greenland

1998). In Study III, we loosen these constraints by using random forests algorithms, which make no assumptions the data distribution or classify the data into a theoretical distribution (Breiman 2001b). However, both strategies rely heavily on identifying and including the correct confounders. Indeed, without randomisation, causal inferences depend on whether we are in possession of a sufficient set of variables for analysis (Pearl, Robins, and Greenland 1999).

A natural question that arises is how to identify the necessary confounders to include in the analyses. The identification of confounders should be made using a priori subject matter knowledge (Hernán, S. Hernández-Díaz, et al. 2002). In Study I, for instance, the association between social capital and mortality is likely to be confounded by health behaviours such as smoking. Individuals with lower social capital smoke more (e.g. G.N. et al. 2011) and smoking causes increases mortality rates (Leistikow et al. 1998). Thus, to not exaggerate the effect of social capital on mortality, we include such health behaviours as control variables in the analysis.

It is important to note that the set of control variables will almost always remain insufficient (Pearl, Robins, and Greenland 1999). Thus, causal inferences from the studies in this dissertation rest on subject-matter judgements about which variables might be missing from the analyses. This is an unavoidable limitation in studies where random assignment is practically impossible. Like all scientific studies, we would like to see our results replicated in different contexts and using other potentially important confounders. When replication studies can confirm our results, our confidence should grow that the mechanisms that we have uncovered are not merely correlational artefacts, but are indeed causal (E. Loken and Gelman 2017; Rothman and Greenland 1998).

4.1.4 The issues of missing observations

Missing data are a statistical problem that occurs when individuals do not respond to one or more items in a survey (Newman 2009). The presence of missing data in a dataset demands that the researcher make a careful choice among several different missing data treatments, i.e. techniques to take the uncertainty caused by missingness into account (Newman 2014). This choice should be made carefully, as choosing an inferior approach may lead to parameter bias and inferential errors

(Sterne et al. 2009). Which techniques that are the most appropriate depend on factors such as the degree of missingness and mechanism that lead to the missing data (Little and Rubin 1987). As a natural consequence of working with different data types from various sources, the missing data problem was approached in different ways in the studies in this dissertation, as I described in Chapter 2.

In Study I, we checked the results robustness to missing data. Multiple imputation by chained equations (MICE) were performed to adjust parameter estimates and standard errors for the possible bias caused by the missing data mechanism. MICE consists of a series of estimations regressing each variable in the model on the other variables. In other words, the procedure loops through the variables predicting each variable dependent on the others. This procedure provides excellent flexibility as each variable may be assigned a suitable distribution (White, Royston, and Wood 2011). For instance, continuous variables are modelled through linear regression, binary variables through logistic regression and so on (Chevret, Seaman, and Resche-Rigon 2015). Several simulation studies have demonstrated that MICE outperforms other techniques at missing rates higher than 10% and various missing data mechanisms and sample sizes (Janssen et al. 2010; Knol et al. 2010).

In Study I, MICE were used in conjunction with a Cox proportional hazards model (Andrea Marshall, Altman, and Holder 2010; Marshall et al. 2009), and only trivial differences from when handling the missing data using listwise deletion were found. Thus, we presented the results based on non-imputed data confident that missing data were of no concern. In Study II, missing data were handled using solely listwise deletion. If the missing data rate is low, there is likely to be little difference between ignoring the missing data and applying more advanced techniques. As the partial missing data rate in Study II was below 10%, my co-authors and I opted for listwise deletion as more advanced techniques were unlikely to produce different estimates (Newman 2014; Wulff and Ejlskov 2017). Here, it also worth noting that more advanced procedures such as MICE become increasingly computationally intensive when the size of the dataset grows (Allison 2002; Bodner 2008). Given the immense sample size in Study II, I saw little reason to employ techniques that were excessively computationally intensive while bringing little value to the analysis.

In Studies III and IV, missing data rates suggested that more sophisticated approaches were called for. In these studies, however, a more recent development in missing data imputation: random forest imputation (RFI) were applied. As described previously in more detail, random forests combine several regression trees into an ensemble of trees each time varying the sample size and the predictors (Hastie, Tibshirani, and Friedman 2009). Random forests are excellent in situations where the imputation model is difficult to specify due to, e.g. complex nonlinear relationships such as interactions (Doove, Van Buuren, and Dusseldorp 2014; Shah et al. 2014). The complexity made RFI an obvious candidate for Studies III and IV where we needed to handle complex interrelationships between many covariates. RFI does not require the researcher to specify an imputation model where important moderating relationships are specified. As one of the primary goals of Study III was to discover the most central of these interactions among the included variables, there was little basis on which to specify an imputation model a priori. This made RFI a desirable alternative to handle missing data in these studies.

4.2 Conceptualization and measurement of social relationships

One of the challenges in this dissertation was how to assess and compare the different measurements and conceptualisations of social relationships. One challenge is that comparing different measurements makes a comparison of effects difficult across studies (Valtorta et al. 2016). The comparison is made further difficult because I have used data that span two countries and the use of different health outcomes. Differences in norms, culture and interpretation may affect what a particular measure represents in the two countries. To get a better handle on how these different conceptualisations may represent similar underlying dimensions of social relationships, I made an overview of which underlying social relationship features these different aspects tap into (see Table 1.2 on page 1.2). However, an in-depth theoretical exploration and comparison are beyond the scope of this dissertation. The primary purpose was to develop a useful overview based on previous literature that may help in seeing the similarities between seemingly different measures of social relationships. Thus, the overview should be viewed as inspirational only and not as a theoretically valid model.

Another challenge is to assess the reliability and validity of each of the used measures. It is beyond the scope of this dissertation to give a detailed overview of the extent to which all of the used measures across of four studies are validated. Instead, I refer to excellent studies such as Valtorta (2016) and Wang et al. (2017). In Studies III and IV the loneliness outcome was measured by the 3-item short UCLA scale (Hughes et al. 2004). Since the validity of this measure had not yet been assessed in a British setting as part of Study III, we used a similar approach to Uysal-Bozkir et al. (2017) to assess the validity of this measure. Overall there was evidence of both convergent, discriminant and structural validity and loneliness levels were very similar to the US-based validation study (Hughes et al. 2004). The assessment is described in more detail in the supplementary materials.

One may criticise the quality of socioeconomic position as a proxy for social relationships in early-life. As argued in the introduction, the proxy in part includes aspects of socialisation and behaviour. However, the proxy also represents a range of other factors such as material and cultural circumstances that do not reflect the social relationships that are a focus of this dissertation.

In summary, comparison of effect sizes and associations should be made with caution across the different studies (I-IV).

4.3 Discussion of the different statistical methods

In this section, I discuss the statistical methodologies used in the studies in this dissertation. I frame this discussion in the context of Breiman's (2001) division of statistical modelling into two cultures: Data and algorithmic modelling. The former is well-known by most researchers and not only preferred in the statistical community but also in the biometrics community (Boulesteix and Schmid 2014). The latter is used very sparsely in most fields and commonly referred to as machine learning. Most researchers are not aware of this perspective, and medical statisticians often exhibit scepticism towards its use (Kruppa, Liu, Biau, et al. 2014; Kruppa, Liu, Diener, et al. 2014). Below, I discuss the methods I apply in this dissertation framed in the context of these two perspectives on statistical modelling.

4.3.1 Data models vs algorithmic models

The statistical methods applied in this dissertation fall into two broad categories reflecting two different cultures in statistical modelling: The data modelling culture and the algorithmic modelling culture (Breiman 2001b). In the data modelling culture, a stochastic data model is assumed that connects the exposure variables to the outcome. Such a model could be, e.g. a linear regression model as in Study IV or a Cox regression model as in Studies I and II. Using these models, we estimate parameters – in the case of this dissertation coefficients - from the data and then use these as information about the relationship between the exposures and the outcome. In contrast, in the algorithmic modelling culture, we search for some function in the form of an algorithm to operate on the predictor variables to predict the outcome of interest.

Data models

When using data models, we researchers believe that we come up with a parametric model that is a good description of the complex mechanism that connects the inputs and output. For instance, in Study I, my co-authors and I argue that the hazard function in the Cox model (Cox 1972) appropriately models the connection between social capital and survival. This is advantageous because we can use hazard ratios to interpret the results and thus explain how social capital is related to survival in the long run. This interpretability is preferable when (causal) explanation is desirable (Boulesteix and Schmid 2014). Put differently; such an approach is useful when the interest is in understanding if and how having more social capital cause people to live longer. The emphasis is put on understanding, as the specification of a data model allows the linking of exposures (e.g. social capital) to outcomes (e.g. survival) in a straightforward way that is understandable (Shmueli 2010). Under certain conditions, the specified model can inform about highly complicated mechanisms in an easy-to-interpret way. For instance, in Study II, the interest is in understanding how adult SEP mediates the relationship between early-life SEP and COPD. This interest makes us specify a data model that accounts for this mediating relationship. We then investigate what this model tells about the presence and size of highly complex mediation effects through time.

Data models are not without drawbacks, however. Imposing parametric or semi-parametric models on complex systems result in lower accuracy and information loss. The problem here is the belief that we can select the same data model that nature used to generate the data. Breiman (2001, p. 204) refers to this as ”an a priori straight jacket” that prevents researchers from dealing with many otherwise interesting statistical problems. We might be able to select the right model, but then again we might not be. This problem is also referred to model misspecification. Model misspecification happens when the model is to restricted. For instance, Cox regression as used in Studies I and II assumes proportional hazards. While there are multiple ways of testing whether this assumption is reasonable, in many cases it is often too restrictive (Schmid, Kestler, and Potapov 2015). Algorithmic models avoid misspecification by not imposing the a priori straight jacket by allowing flexibility. While these models remedy the drawbacks of traditional models, they are not a statistical panacea and have – as I discuss below - their own limitations.

Algorithmic models

When using algorithmic models, the approach is that the data are produced in a black box (Hastie, Tibshirani, and Friedman 2009). The insides of this black box are complex and partly unknowable to the researcher (Breiman 2001a). Instead of assuming a data model, the goal is to find an algorithm that is a good predictor of the outcome when running future predictors through it. Thus, the misspecification problem is elegantly solved by not specifying any model at all. In other words, the focus is shifted from explanation to prediction. This approach is unfamiliar to most researchers in most fields (Strobl, Malley, and Tutz 2009) that are focused on understanding and explaining. However, as Breiman (2001, p. 208) puts it ”[u]sing complex predictors may be unpleasant, but the soundest path is to go for predictive accuracy first, then try to understand why.”

The Rashomon Effect The Rashomon Effect occurs when different individuals having witnessed the same event provide contradictory interpretations (Heider 1988). In statistical modelling, we can think of this effect occurring when many different data models result in about the same minimum error rate. Unfortunately, this is often the case (Breiman

2001a). In loneliness research, there is often a vast number of variables to choose from (Victor, Scambler, et al. 2000). Assuming that there are 30 loneliness predictors to choose from and we wish to form the best linear regression with eight predictors. If we were to consider all subsets, there would be almost 6 million regressions with eight predictors. If we use a criterion such as residual sum-of-squares and pick the regression equation with the lowest, we run into a problem: Many 8-variable equations have a residual sum of squares within 1% of the best performing model (Breiman 1996). Now, what if these models that perform almost equally well state something vastly different about the relationship between e.g. the relationship between religious beliefs and loneliness? Which model do we choose?

Dimensionality and Parsimony

The subset example above leads to the problem of dimensionality. What if we simply have too many loneliness predictors? Traditional methods focus on a subset of predictors and delete the rest. Thus, dimensionality is regarded as a problem that we solve by keeping only the ones with the most information. However, getting this information may be tricky if it is hidden in combinations of the predictors. For instance, in Study IV, we increase the flexibility of our linear regression model by including interaction terms between the earlier social adversity and 1) frequency of contact with friends, 2) frequency of contact with family outside the household and 3) quality of relationship with the closest confidante, respectively.

The relation to the Rashomon Effect is clear: We as researchers have to choose which interactions to specify, thus increasing the ways we restrict the data (Simmons, Nelson, and Simonsohn 2011). The use of interactions in Study IV allows us to explain and understand interactive relationships. However, it also complicates the interpretation of the model, which leads us to the principle known as Occam's Razor or the parsimony principle. The parsimony principle states that more straightforward explanations should be preferred. However, when we use a model to predict, accuracy and simplicity work against each other (Burnham and Anderson 1998). For instance, using a linear regression approach, it would be relatively simple to interpret the relationship between say social isolation and loneliness. This regression model's accuracy would be dwarfed by that of a random forest. The tree ensemble produced by the random forest algorithm, however, would be much more difficult to interpret and to understand. This example illustrates the Occam dilemma when choosing between traditional modelling and algorithmic modelling. Accuracy requires more complicated methods, but interpretable functions are often poorer predictors (Domingos 1999).

4.3.2 Benefits of algorithmic modelling

If we are willing to give up interpretability, an algorithmic approach offers us a way to deal with these problems. In Study III, my co-authors show how aggregating over a broad set of competing decision trees (random forests) can help us solve the problem of which variables to include in our model. The random forest algorithm "grows" successive decision trees by using a random element when each tree is constructed (James et al. 2014). At each node, several of the possible loneliness predictors are choose at random and used to make the best possible split. Thus, many different trees with different loneliness predictors are generated and aggregated. This essentially allows for variables to have different effects in different trees, thus preventing the Rashomon Effect from occurring. In Study III, my co-authors and I show how random forests can solve the variable selection problem by producing variable importance plots identifying the most critical correlates of loneliness. These plots can help us identify relevant variables, which would have been overlooked using traditional methods because we simply did not specify our model correctly.

Random forests are also excellent at handling dimensionality. Interestingly, the random forest algorithm changes dimensionality from a problem to a blessing (Breiman 2001a). That is if we have more predictors we have more information. Many of the vast numbers of correlates of loneliness may contain small pieces of information that are lost if they are deleted. As we show in Study III, researchers can use random forests to extract and combine these pieces of information from many different loneliness correlates. It is important to realise that we do not have to worry about specifying important interactions, as random forests take these into account. As I have described previously, the randomly constructed trees make sure that no fundamental interactions are missed along the way. This means that random forests can help loneliness and social epidemiology researchers discover important aspects of their data that traditional methods cannot. This makes random forests ideal for exploration that may lead to the construction of new theories.

In social epidemiology, we work with highly complex phenomena described by noisy and often large datasets. Sometimes our data may be best described by a data model, but other times an algorithmic approach might be far better. In this dissertation, I have tried taking a broader perspective on statistical methodology than what is traditionally used in the field. My perspective is that we first need to ask how to solve the problem. This does not always mean jumping straight to generating a data model that may lead us unknowingly to questionable conclusions.

Chapter 5

Discussion of results

"My interest is not data, it's the world. [...] The idea is to go from numbers to information to understanding."

– Hans Rosling[1]

In this chapter, I discuss the findings from the four studies (I-IV) in light of the three overall aims of the dissertation. All studies added new knowledge to the gaps described in the introduction. For each aim of the dissertation, the findings will be discussed. First, I discuss the results in relation to the life course models I presented in Chapter 1 (section 5.1). I further discuss the potential buffering effect of social relationships throughout the life course. In section 5.2, I discuss how recursive partitioning - an algorithmic model approach - may help to gain new knowledge within the study of social relationships and health. Lastly, in section (5.3), I discuss the evidence for potential differing effects depending on gender.

[1] Iversen 2017

The overall aim of this dissertation is to contribute with new knowledge on the importance of social relationships for health. I have three sub-aims of this dissertation that will guide the discussion: 1) studying social relationships from a life course perspective, 2) incorporating methodological innovations into the studies of the impact of social relationships on health and 3) investigate potential gender differences in the association between social relationships and health. While the studies used different measures of social relationships, I argued in table 1.1 on page (1.2.1), that the social relationship measures tapped into some of the same underlying features of social relationships. Thus, in the subsequent discussion, I integrate this model into the discussion of the findings.

5.1 Social relationships and health in a life course perspective

Studies II and IV took two different conceptual approaches to the study of social relationships' association with health using a life course perspective. Study II was based on Danish register-based cohort and Study IV on the longest running birth cohort study in the world from the UK. Even though two different conceptualisations of social relationships were used, similar findings was suggested in their associations with their respective health outcomes. In addition, both studies added to the knowledge of how social relationships influence health over the life course.

First, Study IV extends the knowledge of how experiences of social relationship adversity throughout the life course are related to loneliness levels later in life. Childhood, mid-adulthood and later adulthood social relationship adversities are associated with loneliness at age 68. These associations are independently associated even when taking social relationships in later stages of the life course into account. Similarly, in Study II, all markers of adverse SEP are associated with an increase in the risk of developing the COPD, CVD and diabetes from age 30 until mid-adulthood. The analyses suggest a clear social gradient in the size of the estimated associations. Having grown up in a family environment which was located high on all three socioeconomic markers showed additional protective effects compared to those who did not.

Thus, the findings from both Study II and IV suggest that childhood social conditions are associated with different health outcomes.

It is, however, important to note, that the contributions of the two studies - while similar - have important differences. In the introduction, I argued that social relationship adversities represent micro psychosocial mechanisms that are located downstream towards health. In contrast, socioeconomic position represents a higher located upstream factor - that in part determines the factors located downstream. Further, I argued that social relationship adversities tap into different underlying features of social relationships (quality and direct social experiences) compared to early-life SEP. I have argued that SEP is a marker tapping into structural properties of the social network structure and in part indirectly conditions social experiences.

In this dissertation, I focus on two life course models; the sensitive life course model and the accumulation of risk model. Disentangling the life course processes of these model have shown to be very difficult (Hallqvist et al. 2004). For this reason, the subsequent discussion should be viewed as a possible interpretation.

5.1.1 The sensitive life course model

In Figure 1.3 on page 20 I presented two sensitive period models; 1) direct associations between social relationships in childhood with or without independent effects of later life risk factors and 2) direct associations with moderating effects of other social relationships. In this section, I first discuss the findings in relations to the first sensitive period model. This life course model suggests that social relationships lived under in childhood influences adult health independent of adult social conditions (Mishra, Cooper, and Kuh 2010). Subsequently, I discuss the findings in relation to whether there are empirical evidence for moderating effects of other social relationships on the direct association.

Both studies suggest that childhood/early-life may be a sensitive period for later health both regarding physical health (COPD, CVD and diabetes) and mental health (loneliness). Both the social relationship exposure in Study II and IV had associations independent of included confounders and later life social relationship factors. However, as illustrated in Figure 1.4 on page 23, the two studies gave evidence of sensitive periods for health outcomes at different stages of the life course (early to mid-adulthood for Study II and later adulthood for Study IV). Further, for both studies, the results suggested that the closer in time the social relationship exposure was located in relation to the studied health outcome, the stronger the effects was on health.

Our results are largely consistent with previous literature for both SEP (Lawlor, Ebrahim, and George Davey Smith 2005; de Sousa Andrade et al. 2017; Galobardes, Lynch, and George Davey Smith 2008) and loneliness (Peters and Liefbroer 1997; Dykstra and de Jong Gierveld 2004). However, a much greater amount of research has been conducted on the influence of early-life SEP conditions on adult health compared to specific social relationship experiences and loneliness in later life. Thus, the contribution of the findings from the two studies to the sensitive life course model - while similar - have important differences. First, the literature establishing the association between early-life SEP and adult health is very consistent across studies and countries (e.g. Psaltopoulou et al. 2017; Becher et al. 2016; Nandi et al. 2012; Galobardes, Lynch, and George Davey Smith 2007). Thus, the primary aim of Study II was not to investigate the consistency of the direct association between early life SEP and adult health. Rather, the aim was to investigate the precise degree to which the association was mediated by the corresponding adult SEP.

In contrast to Study II, only a few studies have investigated the life course influence of social relationship adversity on loneliness as we did in Study IV. Further, no one with the scope of social relationship adversities measured in this study. Thus, the primary aim here was to investigate evidence for or against life course influences of social relationships on loneliness. That is, the distal and not the more proximal pathways in which these experiences may influence health.

Some support for the findings from Study IV can be found in the literature. Dykstra and de Jong Gierveld (2004) investigated the influence of marital history and risk of loneliness later on and found that divorce heightened the risk of loneliness. Additionally, similar to Peters and Liefbroer's study (1997) of marital dissolutions, we found that more recent social relationship adversities showed a stronger association with loneliness compared to less recent ones. This was also the case for the investigation of early-life and adulthood SEP on lifestyle-related diseases and consistent with previous studies. Thus, both the effect of more proximal and distal social relationship exposures seem to diminish over time in their influence on health.

The buffering effect of earlier social experiences

As part of the analyses in Study IV, we investigated whether earlier social experiences influence the association between current quality and quantity of social relationships and loneliness in later life. That is, whether there is evidence for a moderating effect of earlier experiences on later ones (see Figure 1.3 on page 20). A key finding in Study IV is that the extent to which lack of social contact exacerbates loneliness depends on social experiences in earlier life stages. Further, high current quality of relationships may mitigate the negative effect that earlier social relationship adversities exert on loneliness. Thus, there is support for the sensitive period model which includes effects of social relationship modifiers at different life stages.

In the introduction, I argued that social relationship adversities are direct social experiences that reflect the quality of the social network. In the context of this study, we may view social relationship adversities as a form of social wounds that may continue to exert influence on how one evaluates social relationship in later life stages. In turn, this will influence the degree to which social relationships are protective of adverse health outcomes. In line with the discussion in the previous section about the dark side of social capital (Study I), the results from Study IV might suggest a more general dark side of social relationships. However, the moderating effects also indicate how high quality of current social relationship may buffer the effect of earlier adverse experiences thus pointing to how the ’social wounds’ may be mitigated in a health perspective.

In the introduction, I defined social support as the belief that one is cared for, loved, esteemed and valued. The analyses suggested that the negative effect of the number of social relationship adversities in earlier life stages is diminished for those participants with a high current social relationship quality. This is supported by the literature where it has been suggested that supportive relationships in adulthood may protect against adverse effects of childhood adversities on adult health (Umberson, Williams, et al. 2014). However, the darker side of social relationships is also evident in that the number of social relationship adversities exacerbates the association between current lack of social contact and loneliness at age 68.

An important note is the effects of lasting strain seem to diminish over time. This attenuation of effects as the adverse social relationships grow more distant in time have several implications. It may suggest how much earlier social relationship adversities exacerbate the influence of current social isolation on loneliness. This is in line with the primary findings from Study IV that found more recent social relationship adversities more strongly related to loneliness. Additionally, the study from Peters and Liefbroer (1997) also found evidence of time dilution.

5.1.2 The accumulated life course model

In the introduction, I also presented the accumulation of risk model. In this model, the social relationship risk factor could either accumulate independently, with an additive effect or through chains of risk where one harmful exposure leads to another and then another (Ben-Shlomo and Kuh 2002). Further, the social relationship exposures at different life stages could be either correlated or uncorrelated with each other. The chain of risk model suggests the influence of earlier social relationships on later health that happens through the transmission of risk with adult social relationships accounting for the immediate health impact (Mosquera et al. 2017). The study design of Study IV is ill-suited to assess the evidence for this life course model reliably. Still, it can be argued that the slightly attenuated association of earlier relationship adversities when taking later into account may indicate some evidence that earlier experiences increase the risk of later. However, looking at the correlations between the number of social relationship adversities at the three life stages does not support the hypothesis that a high number of adverse experiences in childhood heightens the risk of a high number in adulthood. In this study, we found evidence of at most a low to moderate correlations between social relationships measures at the different life stages. Only the number of social relationship adversity experiences in mid-adulthood was positively correlated with social relationship adversity experiences in later adulthood. These results are in contrast with previous studies that have shown that early social relationship adversities may launch chains of disadvantages in social relationships over the life course (Umberson, Williams, et al. 2014; Katz et al. 2011; Caspi et al.

2006; Dykstra and de Jong Gierveld 2004).

The discrepancy between previous studies and the results of Study IV may be due to several reasons. 1) In Study IV we used different and more comprehensive measures of social relationship adversities compared to previous studies. 2) The adversities used to capture adverse social experiences differed at the three life stages. This makes it hard to speculate whether the lack of correlations was due to imprecise measures of an underlying lack of correlation. 3) Due to a lack of measures of social relationship adversities in young adulthood there was a big year-gap between childhood (before age 18) and mid-adulthood. As previously mentioned the estimated effect sizes of the regression analysis indicated that the association with later loneliness levels diminished over time. Thus, the lack of correlation between childhood social adversities and later social adversities may merely reflect that we did not have access to the life stage subsequent to childhood. 4) We only had access to a limited number of childhood relationship adversities. Thus, the inconsistency between our results and previous studies may merely reflect that the available measures of social relationship adversities in childhood in this study were too crude to assess this hypothesis. Thus, further studies may want to assess other aspects of childhood relationship adversities to assess evidence for or against the chain of risk model of social relationship adversities across the life course. The results from Study II did in part support a chain of risk model with approximately half of the effect of early-life SEP going through adulthood.

To investigate a potential accumulating effect of social relationship exposures, I have conducted some additional analyses that have been presented in section 3.2.2 based on the analyses from Study II and IV. For both SEP (Study II) and the number of social relationship adversities (Study IV), there was evidence of an accumulating effect of being in the high-risk groups on the risk of adverse health. Those participants who experienced a high degree of social relationship adversities at all three life stages had the most substantial association with higher loneliness levels compared to the group had experienced the lowest (Study IV). For SEP, the evidence was similar to those being in the lowest SEP group in both early-life and adulthood having the highest hazard of being diagnosed with both COPD, CVD and diabetes. The accumulation of risk model states that factors elevating the risk of disease accumulate gradually over the life course. Focusing on social relationships, these results may suggest that social relationship risk factors accumulate over time.

5.2 Gaining new insight using machine learning techniques?

In the introduction, I highlighted two ways in which recursive partitioning from the field of machine learning may help to gain insight into the study of the influence of social relationships on health. First, it enables assessing the relative importance of a large number of factors in relation to a specific outcome while considering all possible interrelations between them. Second, such an approach enables the possibility to identify which combinations of a large number of factors that may best identify high-risk subgroups in populations (Hastie, Tibshirani, and Friedman 2009; James et al. 2014). In section 4.3.1 on page 67, I discussed the differences and advantages of algorithmic models - in which machine learning is a part - compared to traditional data models. In this section, I will discuss the drawbacks and advantages in relation to the specific results of Study III.

In Study III, my co-authors and I aimed at exploiting the advantages of an algorithmic modelling approach to dig into the complex relationships between correlates of loneliness among older adults. Using a random forests algorithm, we assessed the relative importance of 42 correlates of loneliness. These 42 correlates belong to one of five overall domains; personality characteristics, affective states, demographic characteristics, social relations and health. Across these 42 correlates, well-being, the feeling of mastery over important life outcomes, the identity of the closest confidant and being extrovert were the most important correlates of lower levels of loneliness. Interestingly, for the participants in this study, the identity of their closest confidante was more predictive of loneliness levels than marital status. These results are supported by previous literature where functional aspects of social relationships have been shown to have stronger associations with loneliness compared to structural measures (Hawkley 2015). We also demonstrated how regression trees can be used to locate high-at-risk groups of lonely individuals. The regression tree provided examples of equifinality – how different pathways can lead to similar outcomes in loneliness (Cicchetti and Rogosch 1996) as well as examples of compounding of risk or protection through the interaction of different correlates (Scott, Jackson, and Bergeman 2011).

The results have several implications for loneliness research but also for research areas working with a large number of correlates of specific health outcomes. A study with similar aims to Study III was conducted by Dahlberg and McKee (2014). They investigated models of social and emotional loneliness in older people using demographic, psychological, health and social variables. However, in contrast to Study III, this study utilized a data modelling approach using sequential multiple regression to model 20 included correlates of loneliness to social and emotional loneliness, respectively. This study had similar findings in regards to the overall importance of well-being. However, as discussed previously, a traditional data modelling approach has an upper bound on the number of variables that can be included in a model at the same time and risk of multicollinearity and overfitting. Additionally, as I pointed out in the introduction, all interrelations need to be explicitly stated between all the variables in the model through interaction effects. Considering Dahlberg and McKee's (2014) study, they included 20 variables compared to the 42 in Study III. Thus, a algorithmic modelling approach allow for a greater number of included variables - and some that may be highly associated with each other. In this regard, many variables in both of the studies are associated with each other. While this is no issue for the validity of results in Study III, the shared covariance in the Dahlberg and McKee study may bias both effects sizes and p-values for some of the included variables. Lastly, in Dahlberg and Mckee's study no interactions between the included variables were specified in the data model. Thus, the ranking of covariates in this study does not take into account potential moderating effects. In contrast, the relative importance of included variables have all potential interrelations included in the assessment of importance. Not just traditional 2-, 3- or higher order interactions - but also interactions with the variable in question itself. Thus, in the ranking of the importance of correlates, it is highly likely that an algorithmic modelling approach such as used in Study III provide more valid evidence compared to an data modelling approach. However, while random forests provides high predictive accuracy and valid identification of the most important correlates the algorithm itself is a black box. In contrast, in the study by Dahlberg and Mckee, the model used to provide the specific ranking is known. Though- as discussed previously - the Rashomon Effect may be in play here where some other model these authors could have specified with the same predictors may perform equally well or even better than the one they choose. As discussed in section 4.3.1, random forest can help identify relevant correlates of loneliness that may be overlooked by traditional model simply because of model misspecification. In summary, both traditional and more advanced statistical techniques such as recursive partitioning have their place in the investigation of the relationship between social relationships and health.

5.3 Gender differences

Study I and IV investigated how gender or earlier social experiences might moderate the association between different conceptualisations of social relationships and health. Study II and III also investigated whether the association between their respective conceptualisations of social relationships and health outcomes differed depending on gender and found only limited (Study II) or no evidence (Study III).

For most conceptualisations of social relationships and most outcomes, the studies performed in this dissertation show limited or no evidence that gender buffers the association between social relationship measures and health. Thus, the evidence accumulated from the studies conducted in this dissertation indicates that gender may be a buffering factor for some health outcomes but not others and some dimensions of social relationships and not others. Overall, the results here are in line with the amassed evidence in the literature so far. In this literature, similar mixed findings for whether or not gender is a moderator on the association between social relationships and health has been found (Ayalon, Shiovitz-Ezra, and Palgi 2013; Dykstra and de Jong Gierveld 2004; Pikhartova, Bowling, and Victor 2015; Kawachi and Berkman 2014; Cockerham 2007). The different social relationship conceptualizations used in this dissertation combined with the different investigated outcomes make it hard to speculate whether the mixed findings are indicative of 1) measuring error or modelling issues in some studies, 2) true differential effects depending on which health outcome are measures or 3) true differential effects depending on which social relationship dimension that is investigated.

It is interesting that the evidence for gender differences was found for dimensions of social capital - a measure that taps into the appraisal of emotional and health-enhancing resources existing in the social network when investigating all-cause mortality (Study I). The association of the four studied dimensions of individual social capital (trust, expectations of reciprocity, civic engagement and social network) differed in the strength of their association with all-cause mortality. The analyses showed no evidence that expectations of reciprocity (unreliable results for both genders) and civic engagement were differently associated with all-cause mortality depending on gender. In contrast, for trust and social network the results suggested that men and women were differently affected.

The results from Study I indicate that more frequent social contact with friends and family are associated with a greater risk of dying for men but a lower risk for women. Similarly, higher levels of trust seemed to have a protective effect on all-cause mortality for women while no reliable association was found for men. These results may be empirical support for the theoretical notion of the dark side of social capital. That is, that social capital may have both positive and negative impacts on health depending on gender and the dimension studied (Kobayashi et al. 2014; Diez Roux 2008). However, all other associations from Studies II-IV showed similar directions for men and for women. In these studies, for both genders, higher (positive) levels of the different dimensions of social relationships translated into better health outcomes both in the physical and mental health domain.

As I stated in the introduction, all-cause mortality is the simplest way to summarise differences according to a social relationship exposure. If these mixed findings are indicative of true differential effects depending on the health outcome measured, the evidence here points to that the other health outcomes such as COPD, diabetes and CVD are not the ones that account for why differences in all-cause mortality between men and women are observed when looking at social relationships and health.

Since SEP was used as a proxy of social relationship structure in Study II, we need further studies to investigate whether more direct measures of social relationships substantiates the results found here. In that regard, a study has given evidence that men with higher social support appear to engage in more risky behaviours than women such as heavy drinking and diets with a higher fat intake (Ikeda et al. 2011) and that men, in general, engage in more risky leisure activities compared to women (Hyyppä et al. 2007). Thus, using conceptualisations of social relationship that tap into more direct aspects of the underlying social relationship dimensions may shed more light on how social relationships impact lifestyle-related diseases.

It may be worthwhile to investigate whether measures such as trust would have similar findings for mental health outcomes such as loneliness (Study III and IV). High levels of trust may both reflect low levels of perceived social stress and anxiety (Giordano and Lindstrom 2010; Elstad 1998) and may reduce apprehension about other people's behaviour (J. Cacioppo, Hughes, et al. 2006). However, it may also be that feeling lonely subsequently reduce the levels of trust and that trust becomes the mediating factor between loneliness and premature mortality (J. Cacioppo and S. Cacioppo 2018). This highlights the complicated relationship between not just measures of social relationships, gender and health but how different health outcomes may affect each other. It also highlights the efficacy of using a life course perspective where the timing of exposures can be sorted.

Chapter 6

Conclusions

"In literature and in life we ultimately pursue, not conclusions,
but beginnings."
– Sam Tanenhaus[1]

This dissertation has at its nexus the importance of social relationships for health. Referring to the aims of this dissertation as stated in the first chapter on page 23, this chapter presents the main conclusions.

[1] Tanenhaus 1986

Overall this dissertation suggest that regardless of the used conceptualisations of social relationships, health outcomes or data source - social relationships are associated with both physical and mental health.

In the introduction, I referred to the key social epidemiological argument is that we are not all created equal. An added social epidemiological argument could easily be that not only are we not created equal but the way in which our social relations protect against ill health are not either.

The findings in this dissertation suggest that the degree to which current quantity and quality of social relationships protect against ill health are dependent upon both life course experiences and gender. However, the strength of these associations may be dependent upon the aspect of social relationships used and the specific health outcome investigated. The findings further indicates, that different conceptualisations of social relationships across different scientific disciplines at first glance hamper the comparison of associations with health. However, taking a bird's eye view, many of them provide different means of tapping into the same underlying features of social relationships.

Overall, the findings indicate that simply using one conceptualisation of social relationship or looking at one time point will not adequately represent the importance of social relationships for health. The findings suggest that childhood may be a sensitive period and that the effect accumulates over the life course. Further, a high current quality of social relationship may mitigate the negative effect of earlier social experiences. Similarly, current social isolation may have more detrimental effects on mental health depending on earlier experiences. It seems that the association between early-life socioeconomic position is partly mediated by the position attained in adulthood and that the main and mediated effects are strongest for pulmonary obstructive lung disease compared to diabetes and cardiovascular disease.

Seemingly different social relationship aspects tap into the same underlying features of social relationships. All these aspects may work in tandem across the life course in their influence on health. In this dissertation I also demonstrated how data analytical advances in other fields might further social epidemiology by providing a means to assess the relative importance of many identified predictors and may help develop the design and targeting of health interventions.

Chapter 7

Perspectives for future research

'Knowing what
Thou knowest not
Is in a sense
Omniscience."
– Piet Hein [1]

The studies in this dissertation helped to fill in gaps in the knowledge of how social relationships may influence health. Both over the life course and how they might work in tandem to influence physical and mental health. The findings have implications for which next steps to take in research on the impact of social relationships and health. In this chapter, I suggest topics for future research.

[1]Goodreads 2018

Indeed, with a worldwide rising life expectancy and advances in modern medicine and diagnostics most will go through life with one or more clinical diagnosed diseases without necessarily viewing themselves as being unhealthy. Thus, there is a need for studies that can incorporate the complexity of disease and health. As the results in this dissertation indicated, using SEP as a social relationship marker has some limitations. Thus, a useful focus for future research may be to dive into the different social environments that growing up in a particular SEP brings. This way, we may learn more about how SEP may or may not condition the social environment.

A further analytic consideration is to switch the focus from statistical significance to a focus on effect sizes. In this regard, a Bayesian analysis could be a possible next step.

Bayesian analysis does not concern itself with statistical significance leading to a stronger focus effects sizes. Also, the considerable knowledge already gained from previous studies of social relationships impact on health can be explicitly included in the model as prior information. Similarly, neural networks and machine learning is another avenue that removes the researcher from a focus on often arbitrary significance levels to a focus on deep learning and the interrelationship between included variables in relation to a specific outcome. While these data-driven approaches have their limitations - as I have discussed in this dissertation - they also have many advantages that could help further the knowledge in this field.

Many of the basic associations between social relationship measures has already been established. Thus, a stronger focus on potential buffering or negating effects could be a worthwhile next step as I have shown evidence for in this dissertation. The focus here may both be on current effects and within a life course perspective.

Bibliography

Abbott, S and D Freeth (2008). "Social capital and health: starting to make sense of the role of generalized trust and reciprocity". In: *Journal of health psychology* 13.7, pp. 874–883.

Adler, P.S. and S. Kwon (2002). "Social capital: Prospects for a new concept". In: *Academy of Management Review* 27.1, pp. 17–40.

Aida, Jun et al. (2011). "Assessing the association between all-cause mortality and multiple aspects of individual social capital among the older Japanese". In: *BMC Public Health* 11.1, p. 499.

Allison, P D (2002). *Missing data*. Thousand Oaks: Sage.

Antonucci, Toni C and Hiroko Akiyama (1987). "An examination of sex differences in social support among older men and women". In: *Sex Roles* 17.11-12, pp. 737–749.

Ayalon, L, S Shiovitz-Ezra, and Y Palgi (2013). "Associations of loneliness in older married men and women". In: *Aging & Mental Health* 17.1, pp. 33–39.

Becher, H. et al. (2016). "Socioeconomic Conditions in Childhood, Adolescence, and Adulthood and the Risk of Ischemic Stroke". In: *Stroke* 47, pp. 173–179.

Ben-Shlomo, Yoav and Diana Kuh (2002). "A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives". In: *International Journal of Epidemiology* 31.2, pp. 285–293.

Berkman, Lisa, T. Glass, et al. (2000). "From social integration to health: Durkheim in the new millennium". In: *Social Science & Medicine* 51.6, pp. 843–857.

Berkman, Lisa, Ichiro Kawachi, and Maria Glymour (2014). *Social epidemiology*. Ed. by LF Berkman, I Kawachi, and M Glymour. 2nd ed. Oxford: Oxford University Press.

Berkman, Lisa and A. Krishna (2014). “Social Network Epidemiology”. In: *Social Epidemiology*. Ed. by Lisa F Berkman, Ichiro Kawachi, and Maria Glymour. 2nd ed. New York: Oxford University Press. Chap. 14, pp. 234–289.

Bodner, Todd E. (2008). “What Improves with Increased Missing Data Imputations?” en. In: *Structural Equation Modeling: A Multidisciplinary Journal* 15.4, pp. 651–675.

Boulesteix, Anne-Laure and M. Schmid (2014). “Machine learning versus statistical modeling”. In: *Biometrical Journal* 56.4, pp. 588–593.

Bourdieu, P. (1986). “The forms of capital”. In: *Handbook of theory and research for the sociology of education* 241, p. 258.

Bourdieu, P and LJD Wacquant (1992). “An invitation to reflexive sociology”. In: University of Chicago press.

Bowlby, J. (1979). *The making and breaking of affectional bonds*. London: Tavistock Publications Lts.

Box, G.E.P. (1979). “Robustness in the strategy of scientific model building”. In: *Robustness in Statistics*. Ed. by R. L. Launer and G. N. Wilkinson. Academic Press, pp. 201–236.

Breiman, Leo (1996). “Heuristics of instability and stabilization in model selection”. In: *The Annals of Statistics* 24.6, pp. 2350–2383.

— (2001a). “Random forests”. In: *Machine Learning* 45.1, pp. 5–32.

— (2001b). “Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author)”. In: *Statistical Science* 16.3, pp. 199–231. arXiv: 0010.

Brooks, Kathryn P. and Christine Dunkel Schetter (2011). “Social Negativity and Health: Conceptual and Measurement Issues”. In: *Social and Personality Psychology Compass* 5.11, pp. 904–918.

Bureau, Alexandre et al. (2005). “Identifying SNPs predictive of phenotype using random forests.” In: *Genetic Epidemiology* 28.2, pp. 171–82.

Burnham, Kenneth P and David R Anderson (1998). *Model selection and inference: a practical information-theoretical approach*. New York: Springer-Verlag.

Cable, Noriko (2014). “Life course approach in social epidemiology: an overview, application and future implications.” In: *Journal of epidemiology* 24.5, pp. 347–52.

Cacioppo, John and Stephanie Cacioppo (2018). “The growing problem of loneliness.” In: *Lancet* 391.10119, p. 426.

Cacioppo, John, Mary Elizabeth Hughes, et al. (2006). "Loneliness as a specific risk factor for depressive symptoms: cross-sectional and longitudinal analyses." In: *Psychology and Aging* 21.1, p. 140.

Cacioppo, John and W Patrick (2008). *Loneliness: Human nature and the need for social connection*. WW Norton & Company.

Caspi, A et al. (2006). "Socially isolated children 20 years later: risk of cardiovascular disease". In: *Archives of Pediatrics & Adolescent Medicine* 160.8, pp. 805–811.

Cassel, J (1976). "The contribution of the social environment to host resistance: the fourth wade Hampton frost lecture". In: *American Journal of Epidemiology* 104.2, pp. 107–123.

Chevret, S, S Seaman, and M Resche-Rigon (2015). "Multiple imputation: a mature approach to dealing with missing data." In: *Intensive care medicine* 41.2, pp. 348–50.

Cicchetti, D and FA Rogosch (1996). "Equifinality and multifinality in developmental psychopathology". In: *Development and Psychopathology* 8.04, pp. 597–600.

Claussen, B, G Davey Smith, and D Thelle (2003). "Impact of childhood and adulthood socioeconomic position on cause specific mortality: the Oslo Mortality Study." In: *Journal of Epidemiology & Community Health* 57.1, pp. 40–45.

Cobb, Sidney (1976). "Social support as a moderator of life stress." In: *Psychosomatic Medicine* 38.5, pp. 300–314.

Cockerham, W. (2005). "Health Lifestyle Theory and the Convergence of Agency and Structure". In: *Journal of Health and Social Behavior* 46.1, pp. 51–67.

— (2007). *Social causes of health and disease*. Polity.

Conan-Doyle, Arthur (1892). "The Adventure of the Copper Beeches". In: *Adventures of Sherlock Holmes*. Strand Magazine.

Cox, Dr (1972). "Regression Models and Life-Tables". In: *Journal of the Royal Statistical Society.1* 34.2, pp. 187–220.

Dahlberg, Lena and Kevin J. McKee (2014). "Correlates of social and emotional loneliness in older people: evidence from an English community study". In: *Aging & Mental Health* 18.4, pp. 504–514.

De Jong Gierveld, Jenny (1987). "Developing and testing a model of loneliness." In: *Journal of Personality and Social Psychology* 53.1, pp. 119–128.

de Sousa Andrade, Dayse Rodrigues et al. (2017). "Life course socioeconomic adversities and 10-year risk of cardiovascular disease: cross-sectional analysis of the Brazilian Longitudinal Study of Adult Health". In: *International Journal of Public Health* 62.2, pp. 283–292.

Diez Roux, A. V. (2008). *Social Capital and Health*. Ed. by Ichiro Kawachi. Vol. 168. 11. New York, NY : Springer Science + Business Media, LLC, pp. 1340–1342.

Dohrenwend, Bruce P. (2006). "Inventorying stressful life events as risk factors for psychopathology: Toward resolution of the problem of intracategory variability." In: *Psychological Bulletin* 132.3, pp. 477–495.

Domingos, Pedro (1999). "The Role of Occam's Razor in Knowledge Discovery". In: *Data Mining and Knowledge Discovery* 3.4, pp. 409–425.

Donne, John (1923). *Donne's Devotions*. Cambridge University Press.

Doove, L.L., S. Van Buuren, and E. Dusseldorp (2014). "Recursive partitioning for missing data imputation in the presence of interaction effects". In: *Computational Statistics & Data Analysis* 72, pp. 92–104.

Due, P et al. (1999). "Social relations: network, support and relational strain." In: *Social Science & Medicine* 48.5, pp. 661–73.

Dykstra, PA and Jenny de Jong Gierveld (2004). "Gender and marital-history differences in emotional and social loneliness among Dutch older adults". In: *Canadian Journal on Aging/La revue canadienne du vieillissement* 23.2, pp. 141–55.

Ejlskov, Linda, Rikke N Mortensen, et al. (2014). "Individual social capital and survival: a population study with 5-year follow-up." In: *BMC public health* 14.1, p. 1025.

Ejlskov, Linda, Jesper Wulff, et al. (2017). "Assessing the relative importance of correlates of loneliness in later life. Gaining insight using recursive partitioning". In: *Aging & Mental Health*, pp. 1–8.

Elstad, Jon Ivar (1998). "The Psycho-social Perspective on Social Inequalities in Health". In: *Sociology of health & illness* 20.5, pp. 598–618.

Engel, George L. (1977). "The need for a new medical model: a challenge for biomedicine". In: *Science* 196.4286, pp. 129–136.

Fadnes, LT, A Taube, and T Tylleskär (2009). "How to identify information bias due to self-reporting in epidemiological research". In: *The Internet Journal of Epidemiology* 7.2, pp. 1–21.

Galobardes, Bruna, John Lynch, and George Davey Smith (2007). *Measuring socioeconomic position in health research*.

— (2008). "Is the association between childhood socioeconomic circumstances and cause-specific mortality established? Update of a systematic review". In: *Journal of Epidemiology & Community Health* 62.5, pp. 387–390.

Giele, JZ and GH Elder (1998). *Methods of life course research: Qualitative and quantitative approaches*. Ed. by JZ Giele and GH Elder. Sage.

Giordano, Giuseppe N and Martin Lindstrom (2010). "The impact of changes in different aspects of social capital and material conditions on self-rated health over time: a longitudinal cohort study". In: *Social Science and Medicine* 70.5, pp. 700–710.

Glymour, Maria (2007). "Commentary: Selected samples and nebulous measures: some methodological difficulties in life-course epidemiology". In: *International Journal of Epidemiology* 36.3, pp. 566–568.

G.N., Giordano et al. (2011). "The impact of social capital on changes in smoking behaviour: A longitudinal cohort study." In: *European Journal of Public Health* 21.3, pp. 347–354.

Goldberg, Lewis R et al. (1990). "An Alternative " Description of Personality ": The Big-Five Factor Structure". In: *Journal of Personality and Social Psychology* 59.6, pp. 1216–1229.

Goodreads (2018). *Quote by Piet Hein: "Knowing what Thou knowest not Is in a sense Omn..." URL:https://www.goodreads.com/quotes/144776-knowing-what-thou-knowest-not-is-in-a-sense-omniscience*.

Halfon, Neal et al. (2002). "Life course health development: an integrated framework for developing health, policy, and research." In: *The Milbank quarterly* 80.3, pp. 433–479.

Hallqvist, Johan et al. (2004). "Can we disentangle life course processes of accumulation, critical period and social mobility? An analysis of disadvantaged socio-economic positions and myocardial infarction in the Stockholm Heart Epidemiology Program". In: *Social Science & Medicine* 58.8, pp. 1555–1562.

Harpham, T, E Grant, and E Thomas (2002). "Measuring social capital within health surveys: key issues". In: *Health Policy and Planning* 17.1, pp. 106–111.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. Springer.

Hawkley, Louise (2015). “Loneliness and Social Embeddedness in Old Age”. In: *Encyclopedia of Geropsychology*. Singapore: Springer Science+Business Media. Chap. Loneliness.

Hawkley, Louise and John Cacioppo (2010). “Loneliness matters: a theoretical and empirical review of consequences and mechanisms”. In: *Annals of Behavioral Medicine* 40.2, pp. 218–227.

Heider, Karl G. (1988). “The Rashomon Effect: When Ethnographers Disagree”. In: *American Anthropologist* 90.1, pp. 73–81.

Helweg-Larsen, Karin (2011). “The Danish register of causes of death”. In: *Scandinavian Journal of Public Health* 39.7, pp. 26–29.

Hernán, MA, Sonia Hernández-Díaz, and James M. Robins (2004). “A structural approach to selection bias”. In: *Epidemiology* 15.5, pp. 615–625.

Hernán, MA, S. Hernández-Díaz, et al. (2002). “Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology”. In: *American Journal of Epidemiology* 2.155, pp. 176–184.

Holt-Lunstad, Julianne (2018). “Why Social Relationships Are Important for Physical Health: A Systems Approach to Understanding and Modifying Risk and Protection”. In: *Annual Review of Psychology* 69.1, pp. 437–458.

Holt-Lunstad, Julianne, Timothy B Smith, and J Bradley Layton (2010). “Social relationships and mortality risk: a meta-analytic review”. In: *PLoS Medicine* 7.7, e1000316.

Hothorn, Torsten et al. (2006). “Survival ensembles.” In: *Biostatistics (Oxford, England)* 7.3, pp. 355–73.

House, James S, Karl R Landis, and Debra Umberson (1988). *Social relationships and health.*

Hughes, Mary Elizabeth et al. (2004). “A Short Scale for Measuring Loneliness in Large Surveys: Results From Two Population-Based Studies.” In: *Research on Aging* 26.6, pp. 655–672.

Hyyppä, Markku T et al. (2007). “Individual-level measures of social capital as predictors of all-cause and cardiovascular mortality: a population-based prospective study of men and women in Finland”. In: *European Journal of Epidemiology* 22.9, pp. 589–597.

Ikeda, A et al. (2011). “Gender difference in the association between social support and metabolic syndrome in Japan: the ’enkai’ effect?” In: *Journal of Epidemiology and Community Health* 65.1, pp. 71–77.

Iversen, Katja (2017). *7 Data Visualizations that Opened the World's Eyes to Gender Inequality*.
Jahn, Ingeborg et al. (2017). "Examples of sex/gender sensitivity in epidemiological research: Results of an evaluation of original articles published in JECH 2006-2014". In: *Health Research Policy and Systems* 15.1, p. 11.
James, Gareth et al. (2014). *An Introduction to Statistical Learning: with Applications in R*. 6th ed. Springer Publishing Company, Incorporated.
Janssen, Kristel J M et al. (2010). "Missing covariate data in medical research: to impute is better than to ignore." In: *Journal of clinical epidemiology* 63.7, pp. 721–7.
Jensen, Vibeke M. and Astrid W. Rasmussen (2011). "Danish education registers". In: *Scandinavian Journal of Public Health* 39.7_suppl, pp. 91–94.
Jonsson, Ulf et al. (2016). "Psychological Treatment of Depression in People Aged 65 Years and Over: A Systematic Review of Efficacy, Safety, and Cost-Effectiveness". In: *PLOS ONE* 11.8.
Katz, Shaina J et al. (2011). "Childhood social withdrawal, interpersonal impairment, and young adult depression: a mediational model." In: *Journal of Abnormal Child Psychology* 39.8, pp. 1227–38.
Kawachi, Ichiro and Lisa Berkman (2014). "Social Capital, Social Cohesion, and Health". In: *Social Epidemiology*. Ed. by LF Berkman, I Kawachi, and M Glymour. 2nd ed. New York: Oxford University Press. Chap. 8, 290:319.
Kish, Leslie (1959). "Some Statistical Problems in Research Design". In: *American Sociological Review* 24.3, pp. 328–338.
Knol, Mirjam J. et al. (2010). "Unpredictable bias when using the missing indicator method or complete case analysis for missing confounder values: an empirical example". In: *Journal of Clinical Epidemiology* 63.7, pp. 728–736.
Kobayashi, Tomoko et al. (2014). "The Bright Side and Dark Side of Workplace Social Capital: Opposing Effects of Gender on Overweight among Japanese Employees". In: *PloS one* 9.1, e88084.
Krieger, N (2001). "Theories for social epidemiology in the 21st century: an ecosocial perspective". In: *International Journal of Epidemiology*.

Krieger, Nancy (2003). “Genders, sexes, and health: what are the connections—and why does it matter?” In: *International Journal of Epidemiology* 32.4, pp. 652–657.

Kruppa, Jochen, Yufeng Liu, Gérard Biau, et al. (2014). “Probability estimation with machine learning methods for dichotomous and multicategory outcome: Theory”. In: *Biometrical Journal* 56.4, pp. 534–563.

Kruppa, Jochen, Yufeng Liu, Hans-Christian Diener, et al. (2014). “Probability estimation with machine learning methods for dichotomous and multicategory outcome: Applications”. In: *Biometrical Journal* 56.4, pp. 564–583.

Kuh, Diana, Yoav Ben-Shlomo, et al. (2003). “Life course epidemiology.” In: *Journal of Epidemiology and Community Health* 57.10, pp. 778–783.

Kuh, Diana and Rebecca Hardy (2002). *A life course approach to women's health.* Oxford University Press.

Kuh, Diana, Andrew Wong, et al. (2016). “The MRC National Survey of Health and Development reaches age 70: maintaining participation at older ages in a birth cohort study”. In: *European Journal of Epidemiology* 31.11, pp. 1135–1147.

Kushner, Howard I and Claire E Sterk (2005). “The limits of social capital: Durkheim, suicide, and social cohesion.” In: *American Journal of Public Health* 95.7, pp. 1139–43.

Lange, T and JV Hansen (2011). “Direct and indirect effects in a survival context”. In: *Epidemiology* 22.4, pp. 575–581.

Lange, T, S Vansteelandt, and Maarten Bekaert (2012). “A simple unified approach for estimating natural direct and indirect effects”. In: *American Journal of Epidemiology* 176.3, pp. 190–195.

Lange, Theis et al. (2017). “Applied mediation analyses: a review and tutorial”. In: *Epidemiology and Health* 39, e2017035.

Lawlor, Debbie A, Shah Ebrahim, and George Davey Smith (2005). “Adverse socioeconomic position across the lifecourse increases coronary heart disease risk cumulatively: findings from the British women's heart and health study.” In: *Journal of Epidemiology and Community Health* 59.9, pp. 785–93.

Lawrence, Elizabeth M. (2017). “Why Do College Graduates Behave More Healthfully Than Those Who Are Less Educated?” In: *Journal of Health and Social Behavior*, p. 002214651771567.

Leistikow, Bruce N et al. (1998). "Smoking as a risk factor for accident death: a meta-analysis of cohort studies". In: *Preventive Medicine* 27.6, pp. 871–878.

Liaw, A. and M. Wiener (2002). "Classification and Regression by randomForest". In: *R News* 2.3, pp. 18–22.

Lindbladh, E et al. (1997). "The diffusion model and the social-hierarchical process of change". In: *Health Promotion International* 12.4, pp. 323–330.

Link, B G and J Phelan (1995). "Social conditions as fundamental causes of disease." In: *Journal of Health and Social Behavior* Spec No.1995, pp. 80–94.

Little, R. J. A. and Donald B Rubin (1987). *Statistical analysis with missing data*. Vol. 1. New York: John Wiley.

Loken, E and A Gelman (2017). "Measurement error and the replication crisis". In: *Science* 6325.355, pp. 584–585.

Lund, R (2016). "Social relations and health with a special focus on life course and stressful social relations". Doctoral dissertation. University of Copenhagen.

Lund, Rikke et al. (2009). "Exposure to bullying at school and depression in adulthood: A study of Danish men born in 1953". In: *European Journal of Public Health* 19.1, pp. 111–116.

Lunetta, Kathryn L et al. (2004). "Screening large-scale association study data: exploiting interactions using random forests." En. In: *BMC Genetics* 5.1, p. 32.

Macinko, J and B Starfield (2001). "The utility of social capital in research on health determinants". In: *Milbank Quarterly* 79.3, pp. 387–427.

Mann, C J (2003). "Observational research methods. Research design II: cohort, cross sectional, and case-control studies". In: *Emergency Medicine Journal* 20.1, pp. 54–60.

Marshall, Andrea, Douglas G Altman, and Roger L Holder (2010). "Comparison of imputation methods for handling missing covariate data when fitting a Cox proportional hazards model: a resampling study." In: *BMC Medical Research Methodology* 10.1, p. 112.

Marshall, A et al. (2009). "Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines". In: *BMC Medical Research Methodology* 9, p. 57.

Mathers, Colin D and Dejan Loncar (2006). “Projections of Global Mortality and Burden of Disease from 2002 to 2030”. In: *PLoS Medicine* 3.11. Ed. by Jon Samet, e442.

McLeroy, Kenneth R. et al. (1988). “An Ecological Perspective on Health Promotion Programs”. In: *Health Education Quarterly* 15.4, pp. 351–377.

Miller, Candace et al. (2009). “Social Epidemiology”. In: *Social and Behavioral Aspects of Pharmaceutical Care*, pp. 17–35.

Mishra, Gita, Rachel Cooper, and Diana Kuh (2010). “A life course approach to reproductive health: Theory and methods”. In: *Maturitas* 65.2, pp. 92–7.

Mosquera, Paola A et al. (2017). “Are health inequalities rooted in the past? Income inequalities in metabolic syndrome decomposed by childhood conditions.” In: *European Journal of Public Health* 27.2, pp. 223–233.

Murayama, Hiroshi, Yoshinori Fujiwara, and Ichiro Kawachi (2012). “Social capital and health: a review of prospective multilevel studies”. In: *Journal of Epidemiology* 22.3, pp. 179–187.

Nandi, Arijit et al. (2012). “Using marginal structural models to estimate the direct effect of adverse childhood social conditions on onset of heart disease, diabetes, and stroke”. In: *Epidemiology* 23.2, pp. 223–232. arXiv: `NIHMS150003`.

Newman, D A (2009). “Missing data techniques and low response rates: The role of systematic nonresponse parameters”. In: *Statistical and methodological myths and urban legends: Doctrine, verity, and fable in the organizational and social sciences*. Ed. by C E Lance and R J Vandenberg. New York: Routledge, pp. 7–36.

— (2014). “Missing data: Five practical guidelines”. In: *Organizational Research Methods* 17.4, pp. 372–411.

Nyqvist, Fredrica et al. (2013). “Structural and Cognitive Aspects of Social Capital and All-Cause Mortality: A Meta-Analysis of Cohort Studies”. In: *Social Indicators Research*, pp. 1–22.

OECD (2015). *In It Together: Why Less Inequality Benefits All*. Tech. rep. Paris: OECD Publishing.

Pearl, Judea, James M. Robins, and Sander Greenland (1999). “Confounding and Collapsibility in Causal Inference”. In: *Statistical Science* 14.1, pp. 29–46.

Pedersen, Carsten Bøcker (2011). "The Danish civil registration system". In: *Scandinavian Journal of Public Health* 39.7, pp. 22–25.
Peters, Arnold and AC Aart C. Liefbroer (1997). "Beyond marital status: Partner history and well-being in old age". In: *Journal of Marriage and the Family* 59.3, p. 687.
Petersson, Flemming, Mikkel Baadsgaard, and Lau Caspar Thygesen (2011). "Danish registers on personal labour market affiliation". In: *Scandinavian Journal of Public Health* 39.7_suppl, pp. 95–98.
Pikhartova, Jitka, Ann Bowling, and Christina Victor (2015). "Is loneliness in later life a self-fulfilling prophecy?" In: *Aging & Mental Health*, pp. 1–7.
Portes, Alejandro (1998). "Social capital: Its origins and applications in modern sociology". In: *Annual Review of Sociology* 24.1, pp. 1–24.
— (2000). "The two meanings of social capital". In: *Sociological Forum*. Vol. 15. 1. Kluwer Academic Publishers-Plenum Publishers., pp. 1–12.
Pottegård, Anton et al. (2017). "Data Resource Profile: The Danish National Prescription Registry." In: *International journal of epidemiology* 46.3, 798–798f.
Power, Chris, Elina Hyppönen, and George Davey Smith (2005). "Socioeconomic position in childhood and early adult life and risk of mortality: a prospective study of the mothers of the 1958 British birth cohort." In: *American Journal of Public Health* 95.8, pp. 1396–402.
Psaltopoulou, Theodora et al. (2017). "Socioeconomic status and risk factors for cardiovascular disease: Impact of dietary mediators". In: *Hellenic Journal of Cardiology* 58.1, pp. 32–42.
R Core Team, . (2017). *R: A Language and Environment for Statistical Computing*. Vienna.
Region Nordjylland and Statens Institut for Folkesundhed (2013). *Sådan står det til med sundheden i Aalborg Kommune*. Tech. rep.
Ripley, B. (2016). *tree: Classification and Regression Trees*.
Rothman, KJ and Sander Greenland (1998). *Modern Epidemiology*. 2nd ed. Philadelphia: Lippincott-Raven.
Rubin, Donald B (1987). *Multiple imputation for nonresponse in surveys*. New York: John Wiley and Sons.
Savla, J. Tina et al. (2013). "Childhood abuse affects emotional closeness with family in mid- and later life". In: *Child Abuse & Neglect* 37.6, pp. 388–399.

Schmid, M., H. A. Kestler, and S. Potapov (2015). “On the validity of time-dependent AUC estimators”. In: *Briefings in Bioinformatics* 16.1, pp. 153–168.

Schmidt, Morten, Lars Pedersen, and Henrik Toft Sørensen (2014). *The Danish Civil Registration System as a tool in epidemiology*.

Schmidt, Morten, Sigrun Alba Johannesdottir Schmidt, et al. (2015). *The Danish National patient registry: A review of content, data quality, and research potential*.

Schooling, CM (2001). “Health behaviour in a social and temporal context”. PhD thesis. University of London.

Scott, Stacey B, Brenda R Jackson, and C S Bergeman (2011). “What contributes to perceived stress in later life? A recursive partitioning approach”. In: *Psychology and Aging* 26.4, pp. 830–43.

Segal, Mark R, Jason D Barbour, and Robert M Grant (2004). “Relating HIV-1 sequence variation to replication capacity via trees and forests.” In: *Statistical Applications in Genetics and Molecular Biology* 3.

Shah, Anoop D et al. (2014). “Comparison of random forest and parametric imputation models for imputing missing data using MICE: a CALIBER study.” en. In: *American journal of epidemiology* 179.6, pp. 764–74.

Shen, Kai-Quan et al. (2007). “A feature selection method for multilevel mental fatigue EEG classification.” English. In: *IEEE transactions on bio-medical engineering* 54.7, pp. 1231–7.

Shmueli, Galit (2010). “To Explain or to Predict?” In: *Statistical Science* 25.3, pp. 289–310.

Silver, Nate (2012). *The Signal and the Noise: Why So Many Predictions Fail - But Some Don't*. Penguin.

Simmons, Joseph P., Leif D. Nelson, and Uri Simonsohn (2011). “False-Positive Psychology”. In: *Psychological Science* 22.11, pp. 1359–1366.

Singer, J. D. and J. B. Willett (2003). *Applied longitudinal data analysis : modelling change and event occurrence*. Oxford: Oxford University Press.

Stafford, Mai et al. (2013). “Using a birth cohort to study ageing: representativeness and response rates in the National Survey of Health and Development.” In: *European Journal of Ageing* 10.2, pp. 145–157.

Stekhoven, Daniel J (2011). *Using the missForest Package*. Tech. rep., pp. 1–11.

Stekhoven, Daniel J. (2013). *missForest: Nonparametric Missing Value Imputation using Random Forest*.

Stekhoven, Daniel J and Peter Bühlmann (2012). "MissForest–non-parametric missing value imputation for mixed-type data." In: *Bioinformatics (Oxford, England)* 28.1, pp. 112–8.

Steptoe, Andrew et al. (2013). "Social isolation, loneliness, and all-cause mortality in older men and women". In: *Proceedings of the National Academy of Sciences* 110.15, pp. 5797–5801.

Sterne, Jonathan A C et al. (2009). "Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls." English. In: *BMJ (Clinical research ed.)* 338, b2393.

Stringhini, Silvia et al. (2013). "Association of Lifecourse Socioeconomic Status with Chronic Inflammation and Type 2 Diabetes Risk: The Whitehall II Prospective Cohort Study". In: *PLoS Medicine* 10.7. Ed. by Ronald C. W. Ma, e1001479.

Strobl, Carolin, Anne-Laure Boulesteix, Thomas Kneib, et al. (2008). "Conditional variable importance for random forests". En. In: *BMC Bioinformatics* 9.1, p. 307.

Strobl, Carolin, Anne-Laure Boulesteix, Achim Zeileis, et al. (2007). "Bias in random forest variable importance measures: illustrations, sources and a solution". En. In: *BMC Bioinformatics* 8.1, p. 25.

Strobl, Carolin, James Malley, and Gerhard Tutz (2009). "An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests." In: *Psychological methods* 14.4, pp. 323–48.

Susser, Mervyn (1973). *Causal thinking in the health sciences: concepts and strategies of epidemiology*. New York: Oxford University Press.

Tanenhaus, Sam (1986). *Literature Unbound*. Ballantine Books.

Thoits, Peggy A. (2011). "Mechanisms linking social ties and support to physical and mental health". In: *Journal of Health and Social Behavior* 52.2, pp. 145–161.

Thomas, Patricia (2011). "Trajectories of social engagement and limitations in late life". In: *Journal of Health and Social Behavior* 52.4, pp. 430–443.

Thygesen, Lau Caspar and A K Ersbøll (2011). "Danish population-based registers for public health and health-related welfare research–a description of Danish registers and results from their application in research". In: *Scandinavian Journal of Public Health* 39.suppl 7, pp. 8–10.

Uchino, BN, John Cacioppo, and JK Kiecolt-Glaser (1996). "The relationship between social support and physiological processes: a review with emphasis on underlying mechanisms and implications for health." In: *Psychological Bulletin* 119.3, p. 488.

Umberson, Debra, Robert Crosnoe, and Corinne Reczek (2010). "Social Relationships and Health Behavior Across Life Course." In: *Annual review of sociology* 36, pp. 139–157.

Umberson, Debra and Jennifer Karas Montez (2010). "Social relationships and health: a flashpoint for health policy." In: *Journal of Health and Social Behavior* 51 Suppl.Suppl, S54–66.

Umberson, Debra, Kristi Williams, et al. (2014). "Race, gender, and chains of disadvantage: childhood adversity, social relationships, and health." In: *Journal of Health and Social Behavior* 55.1, pp. 20–38.

Uysal-Bozkir, Özgül et al. (2017). "Translation and Validation of the De Jong Gierveld Loneliness Scale Among Older Migrants Living in the Netherlands." In: *The Journals of Gerontology. Series B* 72.1, pp. 109–119.

Valtorta, Nicole K et al. (2016). "Loneliness, social isolation and social relationships: what are we measuring? A novel framework for classifying and comparing tools." In: *BMJ Open* 6.4, e010799.

VanderWeele, Tyler J (2015). *Explanation in causal inference : methods for mediation and interaction*. Oxford University Press.

VanderWeele, Tyler J and Ilya Shpitser (2013). "On the definition of a confounder." In: *Annals of Statistics* 41.1, pp. 196–220.

Victor, Christina, Sasha Scambler, et al. (2000). "Being alone in later life: loneliness, social isolation and living alone". English. In: *Reviews in Clinical Gerontology* 10.04, pp. 407–417.

Victor, Christina and T. Scharf (2005). "Social isolation and loneliness". In: *Understanding quality of life in old age*. Buckingham: Open University Press, pp. 100–116.

Victor, Christina and Keming Yang (2012). "The Prevalence of Loneliness Among Adults : A Case Study of the United Kingdom". In: *The Journal of Psychology* 146.1-2, pp. 85–104.

Wadsworth, Michael and Suzie Butterworth (2006). "Early life". In: *Social Determinants of Health*. Ed. by Michael Marmot and Richard G. Wilkinson. New York. Chap. 3.

Wadsworth, Michael, Diana Kuh, et al. (2006). "Cohort Profile: The 1946 National Birth Cohort (MRC National Survey of Health and Development)". In: *International Journal of Epidemiology* 35.1, pp. 49–54.

Wakefield, Melanie A, Barbara Loken, and Robert C Hornik (2010). "Use of mass media campaigns to change health behaviour." In: *Lancet (London, England)* 376.9748, pp. 1261–71.

Wang, Jingyi et al. (2017). "Social isolation in mental health: a conceptual and methodological review". In: *Social Psychiatry and Psychiatric Epidemiology* 52.12, pp. 1451–1461.

White, Ian R, Patrick Royston, and Angela M Wood (2011). "Multiple imputation using chained equations: Issues and guidance for practice". In: *Statistics in Medicine* 30.4, pp. 377–399.

World Health Organization (2016). *WHO | Constitution of WHO: principles*.

Wulff, Jesper and Linda Ejlskov (2017). "Multiple imputation by chained equations in praxis: Guidelines and review". In: *Electronic Journal of Business Research Methods* 15.1, pp. 41–56.

Zhang, H and B Singer (2013). *Recursive partitioning in the health sciences*. Springer Science & Business Media.

Appendices

Appendix A

Study I

Study I: Ejlskov, L., Mortensen, R. N., Overgaard, C., Christensen, L. R., Vardinghus-Nielsen, H., Kræmer, S. R., Wissenberg, M., Torp-Pedersen, C. & Hansen, C. D. (2014). Individual social capital and survival: a population study with 5-year follow-up. Published in **BMC public health**, 14(1), 1025.

Appendix B

Study II

Study II: Ejlskov, L., Bøggild, H., Hansen, C.D., Wulff, J., Hansen, SM., Strakopf, L.,Lange, T., Gerds, T. & Torp-Pedersen, C. The effect of early-life and adult socioeconomic position on development of lifestyle related diseases. Under review at **European Journal of Public Health.**

Appendix C

Study III

Study III: Ejlskov, L., Wulff, J, Bøggild, H., Kuh, D & Stafford, M. (2017). Assessing the relative importance of correlates of loneliness in later life. Gaining insight using recursive partitioning. Published in **Aging and Mental Health**, 1-8.

Appendix D

Study IV

Study IV: Ejlskov, L., Bøggild, H., Kuh, D. & Stafford, M. Social relationship adversities throughout the life course and risk of loneliness in later life. Revise and Resubmit in **Ageing and Society.**

ISSN (online): 2246-1302
ISBN (online): 978-87-7210-219-1

AALBORG UNIVERSITY PRESS